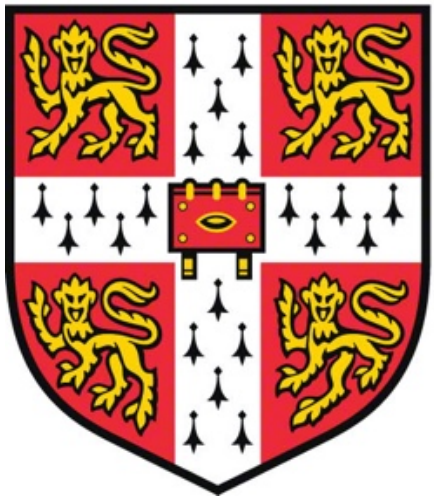
The Utility of CA125 for the Detection of Ovarian Cancer in Primary Care



Dr Garth Funston
Gonville and Caius College

The Department of Public Health and Primary Care
University of Cambridge

This thesis is submitted for the degree of:
Doctor of Philosophy

April 2021

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has been submitted before for any degree or other qualification. It does not exceed the prescribed limit of 60,000 words (excluding figures, tables, the bibliography and appendices) set by the Degree Committee for Clinical Medicine.

Summary

Background

Ovarian cancer is the 6th most common cancer to affect UK women and has the worst prognosis of any gynaecological cancer. Most women are not diagnosed until the disease is advanced, which leads to poor outcomes. Earlier ovarian cancer diagnosis has the potential to improve these outcomes. Cancer antigen 125 (CA125) is recommended by the National Institute for Health and Care Excellence (NICE) as the first line test for ovarian cancer in symptomatic women presenting to primary care in England. However, the performance of CA125 in this setting is unknown. The overarching aim of this thesis was to determine the diagnostic performance of CA125 for the detection of ovarian cancer when used in primary care, and to develop and evaluate novel approaches to improve its performance and clinical utility.

Key methods

I used routinely collected primary care and cancer registry data from 50,780 women who underwent CA125 testing in England between 1st May 2011 – 31st December 2014. First, I performed a diagnostic accuracy study, calculating the performance of CA125 within the cohort, at the national cut-off (≥35 U/ml), for the detection of ovarian cancer. Diagnostic accuracy metrics were also calculated for other types of cancer and all cancer types combined (secondary study outcomes). I used logistic regression to estimate the probability of ovarian cancer at specific CA125 levels (1-1000 U/ml) for women of different ages. CA125 levels equating to a 3% ovarian cancer probability (the "risk threshold" at which NICE advocates urgent specialist cancer investigation) were identified. Next, I examined the associations between CA125 test result and time from testing to diagnosis, tumour type and cancer stage, in those women with ovarian cancer. Finally, I developed and internally validated ovarian cancer diagnostic prediction models (of varying complexity) in a sub-group of women with a relevant symptom recorded prior to CA125 testing (n=29,962). To inform the development of these models, I conducted a systematic review of existing ovarian cancer detection tools.

Key results

CA125 had a sensitivity of 77%, a specificity of 94% and a Positive Predictive Value (PPV) of 10% for ovarian cancer at the national cut-off (≥35 U/ml). The PPV for all cancers combined was 21% overall, and 33% in women ≥50 years of age. 20% of women ≥50 years with a raised CA125 level, but no ovarian cancer, had another type of cancer. A CA125 value of 53 U/ml equated to a 3% probability of ovarian cancer overall, but this varied markedly by age (40-year-old: 104 U/ml, 70-year-old: 32 U/ml). Women with a 'normal' CA125 (<35 U/ml) prior to ovarian cancer diagnosis took twice as long to be diagnosed as those with an 'abnormal' CA125, but more frequently had indolent tumour types and were more likely to be diagnosed at an early stage. An ovarian cancer prediction model, incorporating patient age and CA125 level, outperformed CA125 alone. This model showed excellent discrimination on internal validation (AUC: 0.94). Including symptoms, baseline risk factors and other routine blood tests did not improve model performance.

Conclusions

My findings demonstrate that CA125 is a useful test for ovarian cancer detection in primary care. They also indicate that clinicians should consider other types of cancer in women with high CA125 levels, especially if ovarian cancer has been excluded, in order to prevent diagnostic delay. The models presented in this thesis will allow patients and clinicians to determine the estimated probability of ovarian cancer at any given CA125 level and age. This information could inform individual patient decisions on the need for further investigation. If incorporated into the diagnostic pathway, the models would enable patients to be referred on the basis of ovarian cancer risk rather than a generic CA125 cut-off.

Acknowledgements

I am extremely grateful to my primary supervisor, Fiona Walter, and my secondary supervisors, Willie Hamilton and Emma Crosbie, for their advice and encouragement throughout my PhD. They have helped me develop my research ideas and to gain confidence in my abilities. I feel privileged to have had such a supportive supervisory team.

I would like to thank Gary Abel, who provided expert statistical advice on the research presented in **Chapters 3** and **6.** He kindly shared Stata code which I used in **Chapter 3** to estimate the probability of cancer at specific CA125 levels and ages from the results of my logistic regression analyses. In addition, he provided Stata code which I used to lay out the panels in **Figure 3.7** and code which I used to calculate cross-validation calibration slopes for the models presented in **Chapter 6.** I would also like to express my thanks Luke Mounce, who provided expert statistical advice on the research presented in **Chapter 4** and to Sarah Price who provided advice on Stata coding.

I am very grateful to Brian Rous, from the NCRAS, for his advice on cancer coding systems and cancer staging data. He was generous with his time, reviewing lists of cancer codes and classifications I had prepared prior to their use (**Chapters 2** and **3**).

I would like to acknowledge Marije Van Melle, who acted as a second reviewer for the systematic review presented in **Chapter 1**, screening titles and abstracts as well as extracting data from English language guidelines. I would also like to acknowledge Marie-Louise Ladegaard Baun, Henry Jensen and Charles Helsper who contributed to this review by extracting data from foreign language guidelines. Victoria Hardy acted as the second reviewer for the systematic review presented in **Chapter 5** – thank you for your help. Jon Emery kindly provided advice when I was developing the protocol for the review presented in **Chapter 5** and when I was selecting a risk of bias tool this review. I am very grateful for his advice.

I would like to extend my heartfelt thanks to Margaret Johnson - a tireless and passionate advocate for patients. She provided helpful input at the very start of my PhD during my development of study protocols and applications for regulatory approval.

This thesis would not have been possible without the funding provided by Cancer Research UK, through the CanTest initiative, and the NIHR SPCR. I am very grateful for their support.

One person has stood by my side during every twist and turn of my PhD. Through dinner table chat, midnight walks and conversations during occasional sleepless nights you helped me see the bigger picture and kept me going. Your support means the world to me. Thank you to my fiancé Adam. Lastly, I would like to thank a man who, throughout his life, has inspired many young people to achieve things that they did not imagine themselves capable of achieving. Me among them. Thanks Dad.

Publications, presentations and travelling fellowships

Publications resulting from this thesis

Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. **Garth Funston**, Marije Van Melle, Marie-Louise Ladegaard Baun, Henry Jensen, Charles Helsper, Jon Emery, Emma J. Crosbie, Matthew Thompson, Willie Hamilton & Fiona M. Walter. *BMC Cancer*. 2019; 19:028

- Selected as a *BMC Series Highlight* (November 2019)

The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. **Garth Funston**, Willie Hamilton, Gary Abel, Emma J. Crosbie, Brian Rous, Fiona M. Walter. *PLOS Med*. 2020; 17:e1003295

Identifying ovarian cancer in symptomatic women: a systematic review of clinical tools. **Garth Funston**, Victoria Hardy, Gary Abel, Emma J. Crosbie, Jon Emery, Willie Hamilton, Fiona M. Walter. *Cancers*. 2020; 8:3686

- Selected as a *Most Notable Cancers article* (December 2020 - February 2021 period)

CA125 test result, test-to-diagnosis interval and stage in ovarian cancer at diagnosis: a cohort study using electronic health records. **Garth Funston**, Luke TA Mounce, Sarah Price, Brian Rous, Emma J. Crosbie, Willie Hamilton, Fiona M. Walter. *Br J Gen Pract.* 2021; BJGP.2020.0859 (Online First)

Could ovarian cancer prediction models improve the triage of symptomatic women in primary care? A modelling study using routinely collected data. **Garth Funston**, Gary Abel, Emma J. Crosbie, Willie Hamilton, Fiona M. Walter.

Presentations resulting from this thesis

Oral presentations:

The predictive value of CA125 for ovarian cancer in General Practice: a population-based cohort study of over 50,000 women. Late Breaking Abstract and Proffered Paper, National Cancer Research Institute (NCRI) Conference, Glasgow, 2019.

- Awarded a Conference Prize

The association between CA125 level, diagnostic intervals and stage at diagnosis in ovarian cancer: analysis of CPRD and NCRAS data. Cancer and Primary Care Research International Network (Ca-PRI), Toronto, 2019.

Poster presentations:

Variation in the initial assessment for ovarian cancer in symptomatic women: a systematic review of international guidelines. Cancer Research UK Early Detection Conference, Birmingham 2019.

The incidence of non-ovarian cancers in women with elevated CA125 levels in General Practice. National Cancer Research Institute (NCRI) Conference, Glasgow, 2019.

Variation in the initial assessment for ovarian cancer in symptomatic women: a systematic review of international guidelines. Cancer and Primary Care Research International Network (Ca-PRI), Toronto, 2019.

Invited talks

Detecting ovarian cancer in symptomatic women. World Ovarian Cancer Coalition Symposium on Rapid Diagnosis, Online, 2021.

CA125 test result, test-to-diagnosis interval and stage in ovarian cancer at diagnosis: a cohort study using electronic health records. The British Journal of General Practice - Cancer Webinar, Online, (*Scheduled: 26 May 2021*)

Additional publication contributions made during my doctoral research

Recognising gynaecological cancer in primary care: risk factors, red flags and referrals. **Garth Funston**, Helena O'Flynn, Neil AJ Ryan, Willie Hamilton, Emma J. Crosbie. *Adv Ther*. 2018; 35:577-589

Prostate cancer in primary care. Sam Merriel, **Garth Funston**, Willie Hamilton. *Adv Ther*. 2018; 35:1285-1294

Improving the timely detection of bladder and kidney cancer in primary care. Yin Zhou, **Garth Funston**, Georgios Lyratzopoulos, Fiona M. Walter. *Adv Ther*. 2019; 36:1778-1785

Recognising Skin Cancer in Primary Care. Owain T. Jones, Charindu K.I. Ranmuthu, Per N. Hall, **Garth Funston**, Fiona M. Walter. Adv Ther. 2020; 37:603-616

Identifying Novel Biomarkers Ready for Evaluation in Low-Prevalence Populations for the Early Detection of Upper Gastrointestinal Cancers: A Systematic Review. Natalia Calanzani, Page Druce, Claudia Snudden, Kristi Milley, Rachel Boscott, Dawnya Behiyat, Smiji Saji, Javiera Martinez-Gutierrez, Jasmeen Oberoi, **Garth Funston**, Mike Messenger, Jon Emery, Fiona M. Walter. *Adv Ther*. 2020; 38:793-834

Associations between breast cancer survivorship and adverse mental health outcomes: a matched population-based cohort study in the United Kingdom (1988-2018). Helena Carreira, Rachael Williams, **Garth Funston**, Susannah Stanway, Krishnan Bhaskaran. *PLOS Med*. 2021; 18:e1003504

Time from presentation to pre-diagnostic chest X-ray in symptomatic lung cancer patients: a cohort study using electronic patient records. Kirsten Arendse, Fiona M. Walter, Mark Pilling, Yin Zhou, Willie Hamilton, **Garth Funston**. *Br J Gen Pract.* 2021; 71:e273-e279

Recognising Colorectal Cancer in Primary Care. Natalia Calanzani, Aina Chang, Marije Van Melle, Merel M. Pannebakker, **Garth Funston**, Fiona M. Walter. *Adv Ther*. 2021; doi:10.1007/s12325-021-01726-6 (*Epub ahead of print*)

Identifying Novel Biomarkers Ready for Evaluation in Low-Prevalence Populations for the Early Detection of Lower Gastrointestinal Cancers: A Systematic Review and Meta-Analysis. Page Druce, Natalia Calanzani, Claudia Snudden, Kristi Milley, Rachel Boscott, Dawnya Behiyat, Javiera Martinez-Gutierrez, Smiji Saji, Jasmeen Oberoi, **Garth Funston**, Mike Messenger, Jon Emery, Fiona M. Walter. *Adv Ther*. 2021; doi.org/10.1007/s12325-021-01645-6 (*Epub ahead of print*)

Comparison of Two Immunoassays for the Measurement of Serum HE4 for Ovarian Cancer. Chloe E. Barr, **Garth Funston**, Luke T.A. Mounce, Philip W. Pemberton, Jonathon D. Howe, Emma J. Crosbie. *Pract Lab Med.* 2021; (*In press*)

Under review:

The Impact of COVID-19 on Testicular Cancer Presentations: A Supranetwork Multidisciplinary Team Experience. Kenrick Ng, Florentia Mina, Peter Wilson, Daniel M. Berney, Anju Sahdev, Constantine Alifrangis, Han Wong, Sarah Rudman, Michelle Greenwood, Wendy Ansell, **Garth Funston**, Jonathan Shamash.

Travelling fellowships undertaken during my doctoral research

University of Exeter, UK. hosted by Professor Willie Hamilton. May 2018.

University of Washington, USA. Hosted by Professor Matthew Thompson. June 2018.

University of Melbourne, Australia. Hosted by Professor Jon Emery. November 2019.

Contents

List of tables

List of figures

Abbreviations

ACE program - Accelerate Coordinate Evaluate program

ADNEX model - the Assessment of Different Neoplasias in the Adnexa model

AFT model - Accelerated Failure Time model

AIC - Akaike Information Criterion

AUC - Area Under the (receiver operator characteristic) Curve

BMI - Body mass Index

BRCA1 - Breast Cancer susceptibility gene 1

BRCA2 - Breast Cancer susceptibility gene 2

CA125 - Cancer Antigen 125 *aka Carbohydrate Antigen 125*

CCA - Complete Case Analysis

CG122 - Clinical Guideline 122

CI - Confidence Interval

CIBH - Change In Bowel Habit

CPRD - Clinical Practice Research Datalink

CRP - C-Reactive Protein

CT - Computed Tomography

DOvE study - The Diagnosing Ovarian cancer Early study

eCDS tool - electronic Clinical Decision Support tool

EPV - Events Per Variable

eRAT - electronic Risk Assessment Tool

ESR - Erythrocyte Sedimentation Rate

FIGO - Fédération Internationale de Gynécologie et d'Obstétrique

GI - Gastrointestinal

GLOBCAN - Global Cancer Observatory

GP - General Practitioner

HE4 - Human Epididymis protein 4

HES APC - Hospital Episodic Statistics Admitted Patient Care

HGS tumour - High Grade Serous tumour

IARC - The International Agency for Research on Cancer

IBS - Irritable Bowel Syndrome

ICBP - International Cancer Benchmarking Project

ICD-O2 - International Classification of Diseases for Oncology 2rd edition

ICD-O3 - International Classification of Diseases for Oncology 3rd edition

ICD10 - International Classification of Diseases 10th revision

IKU/L - International Kilounits per Litre

IQR - Interquartile Range

ISAC - Independent Scientific Advisory Committee

IU- International units

KU/L - kilounits per litre

LGS tumour - Low Grade Serous tumour

MAR - Missing At Random

MCAR - Missing Completely At Random

MDC - Multidisciplinary Diagnostic Centre

MHRA - Medicines and Healthcare products Regulatory Agency

MI - Multiple Imputation

MICE - Multivariate Imputation by Chained Equation

ml - millilitre

MNAR - Missing Not At Random

MUC-16 - Mucin 16

NCRAS - National Cancer Registration and Analysis Service

NG12 - NICE Guideline 12

NHS - National Health Service

NICE - National Institute for Health and Care Excellence

NLP - Natural Language Processing

NOS - Not Otherwise Specified

NPV - Negative Predictive Value

OC125 antibody - Ovarian Cancer 125 antibody

OR - Odds Ratio

PLCO cancer screening trial - Prostate, Lung, Colorectal and Ovarian cancer screening trial

PPV - Positive Predictive Value

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analysis

PROBAST - Prediction model Risk Of Bias Assessment Tool

PSA -Prostate Specific Antigen

PV - Plasma Viscosity

QUADAS-2 - Quality Assessment of Diagnostic Accuracy Studies 2

RDC - Rapid Diagnostic Centre

RMI - Risk of Malignancy Index

ROB - Risk Of Bias

ROCA - Risk of Ovarian Cancer Algorithm

ROCkeTS trial - Refining Ovarian Cancer Test accuracy Scores trial

ROMA - Risk of Ovarian Malignancy Algorithm

SGO - Society of Gynaecologic Oncology

SI - Symptom Index

STIC - Serous Tubal Intraepithelial Carcinoma

TNM - Tumour Nodes Metastasis

TR - Time Ratio

U/ml - Units per millilitre

UK - United Kingdom

UKCTOCS - UK Collaborative Trial of Ovarian Cancer Screening

UTS - Up-To-Standard

WCRF - World Cancer Research Fund

WHO - World Health Organisation

Chapter 1. Introduction

1.1 Thesis aim and objectives

The overarching aim of this thesis was to evaluate the diagnostic performance of the serum biomarker Cancer Antigen 125 (CA125) for the detection of ovarian cancer when used in primary care and to develop and evaluate novel approaches to improve its performance and clinical utility. CA125 is first and foremost an ovarian cancer test and I have primarily focused on its role in ovarian cancer detection in this thesis. However, given reports that CA125 can be elevated in other types of cancer, I also sought to examine its accuracy for cancer as a whole (all forms) and to explore its utility in the detection of non-ovarian cancer types when used in primary care.

In line with this aim, I set out a series of specific objectives:

i) To evaluate the diagnostic accuracy of CA125 for a) ovarian cancer, b) non-ovarian cancers and c) all cancers combined, when used in English primary care.

ii) To explore variations in the diagnostic accuracy of CA125 by patient age.

iii) To explore the association between CA125 level and estimated probability of a) ovarian cancer and b) all cancers combined, in women undergoing CA125 testing in English primary care.

iv) To identify the CA125 level at which a 3% probability of a) ovarian cancer and b) all cancers combined is reached in women undergoing CA125 testing in English primary care.

v) To explore variations in estimated cancer probability by age and determine the CA125 level at which a 3% probability of a) ovarian cancer and b) all cancers combined is reached in different ages.

vi) To examine the association of pre-diagnostic primary care CA125 result with time between testing and diagnosis, tumour morphology and disease stage in ovarian cancer.

vii) To perform a systematic review to identify published symptom predicated tools for ovarian cancer detection and to compare these tools in terms of a) included variables and b) diagnostic accuracy.

viii) To develop and internally validate an ovarian cancer diagnostic prediction model incorporating symptoms, test results (including CA125), and risk factors.

ix) To explore the potential diagnostic implications of implementing 'action thresholds', based on prediction model derived estimated ovarian cancer probabilities, within primary care.

1.2 Thesis structure

This thesis consists of 7 chapters. In **Chapter 1 (Sections 1.3-1.10)**, I introduce the topic of ovarian cancer detection and outline the rationale for this thesis. In **Chapter 2**, I detail the data sources used in this thesis and describe the initial preparation of data. In **Chapters 3, 4, 5** and **6** I present four separate research studies, each designed to address specific thesis objectives. These chapters follow the standard format of a research paper with Introduction, Methods, Results, Discussion and Conclusion sections. In **Chapter 7,** I bring together the key research findings, consider their implications for clinical practice, and outline future research plans which build on this thesis.

Chapter 1: Introduction

In **Sections 1.3 – 1.10** of **Chapter 1,** I provide essential background information on ovarian cancer and its diagnosis and set out the rational for this thesis.

As background to this thesis I discuss the epidemiology of ovarian cancer and go on to describe the different types of ovarian cancer that can occur. I then consider current approaches to ovarian cancer detection. This includes the presentation of a systematic review of international guidelines on ovarian cancer detection, which I undertook to inform and direct the rest of my doctoral research. I go on to discuss the CA125 test and its role in cancer detection in detail. I summarise relevant literature and highlight gaps in current understanding of CA125 test performance. I conclude this chapter by drawing on the information presented to set out the rational for the research included within this thesis.

Chapter 2: Data sources and data preparation

To conduct all primary research studies included in this thesis (**Chapters 3, 4** and **6**) I used a single dataset, consisting of data from four routinely collected sources. In this chapter, I provide an overview of these data sources and describe the regulatory approval obtained in order to use them. I go on to define the baseline patient cohort. This directly formed the study cohort for **Chapter 3** and subsets of patients were selected from this baseline cohort for the research presented in **Chapters 4** and **6**. Finally, I describe the steps taken to prepare the baseline cohort and several key variables used across multiple thesis studies.

Chapter 3: The diagnostic performance of CA125 for the detection of cancer in primary care: a population-based cohort study (Objectives i-v)

This chapter focuses on the diagnostic accuracy of CA125 in primary care and explores how the probability of cancer changes with CA125 level and patient age. I present a population-based cohort study, which includes two principle types of analysis. First, I report measures of CA125 diagnostic accuracy, which were calculated after applying the conventional threshold (≥35 U/ml), for ovarian cancer, non-ovarian cancers and all types of cancer. Second, I report the results of logistic regression analyses which were used to estimate the probability of cancer based on CA125 level and age. Based on these analyses, I detail the CA125 levels equating to a 3% probability of cancer – the National Institute for Health and Care Excellence (NICE) 'risk threshold' for urgent cancer referral. To conclude **Chapter 3**, I consider the key study findings within the context of the literature and explore their potential implications. This chapter addresses thesis **Objectives i-v**.

Chapter 4: Associations between primary care CA125 test result with test-diagnosis interval and stage in ovarian cancer (Objective vi)

This chapter focuses on potential reasons for, and implications of, false negative CA125 results in primary care. I present the results of a cohort study in which I explored the association of primary care pre-diagnostic CA125 test results ('normal' or 'abnormal') with time between testing and diagnosis (test-diagnosis interval) and stage at diagnosis in ovarian cancer patients. Within this chapter I also examine the types of ovarian tumour which occur in women with normal and abnormal CA125 results. Thesis **Objective vi** is addressed in this chapter.

Chapter 5: Symptom based tools for ovarian cancer detection: a systematic review (Objective vii)

In this chapter, I present a systematic review designed to address thesis **Objective vii**. I describe the systematic identification of studies which have developed or validated ovarian cancer diagnostic tools containing symptom variables. The variables included in these tools are summarised and the diagnostic accuracy of tools compared. I used the results of this review to directly inform the development of a multivariable prediction model in **Chapter 6**.

Chapter 6: The development and internal validation of ovarian cancer diagnostic prediction models for use in symptomatic women in primary care (Objectives viii and ix)

In this chapter, I describe the development and internal validation of two multivariable diagnostic prediction models for ovarian cancer, derived using routinely collected primary care data. In this work, I incorporated patient variables associated with ovarian cancer risk (identified in **Chapter 6** and from other literature sources) alongside CA125, with the aim improving diagnostic accuracy. I present model diagnostic performance metrics and compare them with those of CA125 alone. In addition, I explore the potential implications of using various model derived thresholds (instead of the conventional CA125 cut-off) on the number of ovarian cancers detected in women undergoing CA125 testing in English primary care. Thesis objectives **viii** and **ix** are addressed within this chapter.

Chapter 7: Discussion

In the final chapter, I summarise my research findings and consider how they could help improve cancer diagnostic pathways. I discuss planned research, intended to build on this thesis and address remaining unanswered questions.

1.3 Ovarian cancer epidemiology

1.3.1 Ovarian cancer incidence

In 2020, an estimated 313,959 women were diagnosed with ovarian cancer worldwide.[1] The 2020 Global Cancer Observatory (GLOBOCAN) project estimated that ovarian cancer accounted for 3.4% of new female cancer cases.[1] It is the 6th most commonly diagnosed type of cancer in women in the United Kingdom (UK) with 7,309 new diagnoses made in 2017.[2] In that year, the age standardised ovarian cancer incidence rate in the UK was 22 per 100,000 females.[2*] There has been a gradual decline in ovarian cancer incidence in recent years in high income countries.[3] In the UK, the age standardised incidence rate was around 5% lower during 2015-2017 than 2005-2007.[2] The reasons for this decline are not fully understood, but a reduction in the use of hormone replacement therapy (a risk factor) and the long term effects of higher oral contraceptive medication uptake (a protective factor) are likely to have contributed.[4]

Ovarian cancer predominantly affects older women (**Figure 1.1**). Between 2015-2017, more than 80% of UK diagnoses were made in women aged 50 years or older and more than a quarter in women aged 75 years or older.[5] Incidence rates increase with age, peaking in women aged 75-79 years (72 per 100,000 women), before declining.

Differences in incidence rates have been noted between ethnic groups in several countries, including the United States and the UK.[6,7] The most recent comparison from England (which utilised cancer registry data from 2002-2006) reported that age standardised incidence rates were higher in white women than in other ethnic groups (Black: 7-12 per 100,000; Asian: 9-16 per 100,000; White: 17-18 per 100,00).[7†]

[*] Age standardisation performed using European Standard Population data.[337]
[†] Ranges provided reflect different ways of handling missing ethnicity data in the study analysis.

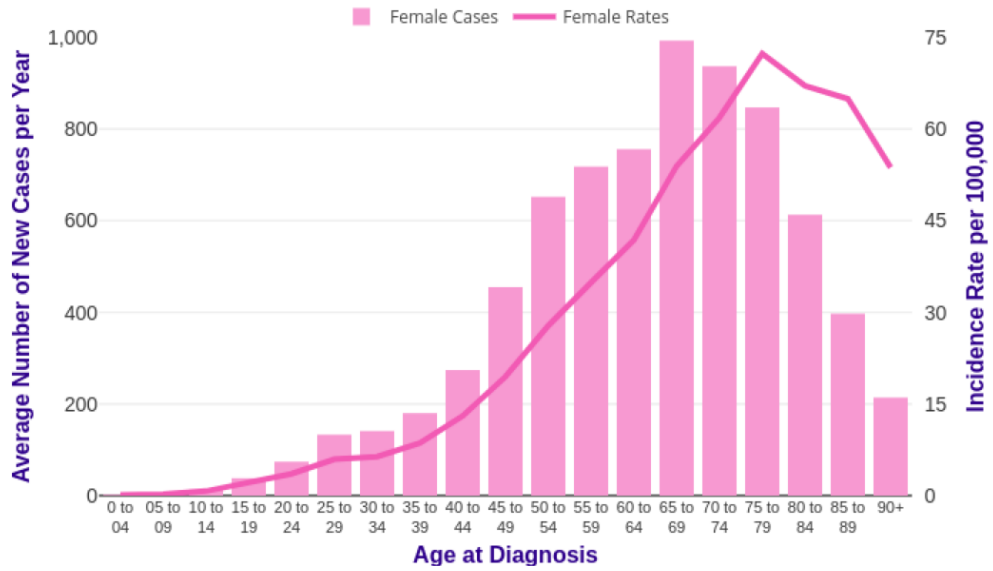


Figure 1.1. Ovarian cancer (C56-C57.4), Average Number of New Cases per Year and Age-Specific Incidence Rates per 100,000 Females, UK, 2015-2017.
Source: CRUK.org/cancerstats[5]

1.3.2 Stage at diagnosis

The majority of women with ovarian cancer are diagnosed at an advanced stage. Of women diagnosed in England between 2010-2014, for whom staging data was available, 42% had stage I or II disease (often considered collectively as 'early stage') and 58% had stage III or IV disease (often considered collectively as 'late stage') (**Figure 1.2**).[8] Recent data from the International Cancer Benchmarking Project (ICBP) found only minor differences in ovarian cancer stage distribution between seven high income countries including the UK.[9‡] A greater proportion of ovarian cancers are detected at a late stage in the UK than other types of gynaecological cancer such as endometrial and cervical cancer.[10,11]

‡ ICBP applied a different definition of ovarian cancer to the one used in this thesis (detailed in in **Section 1.4.4**) as they excluded Borderline tumours from their analyses.

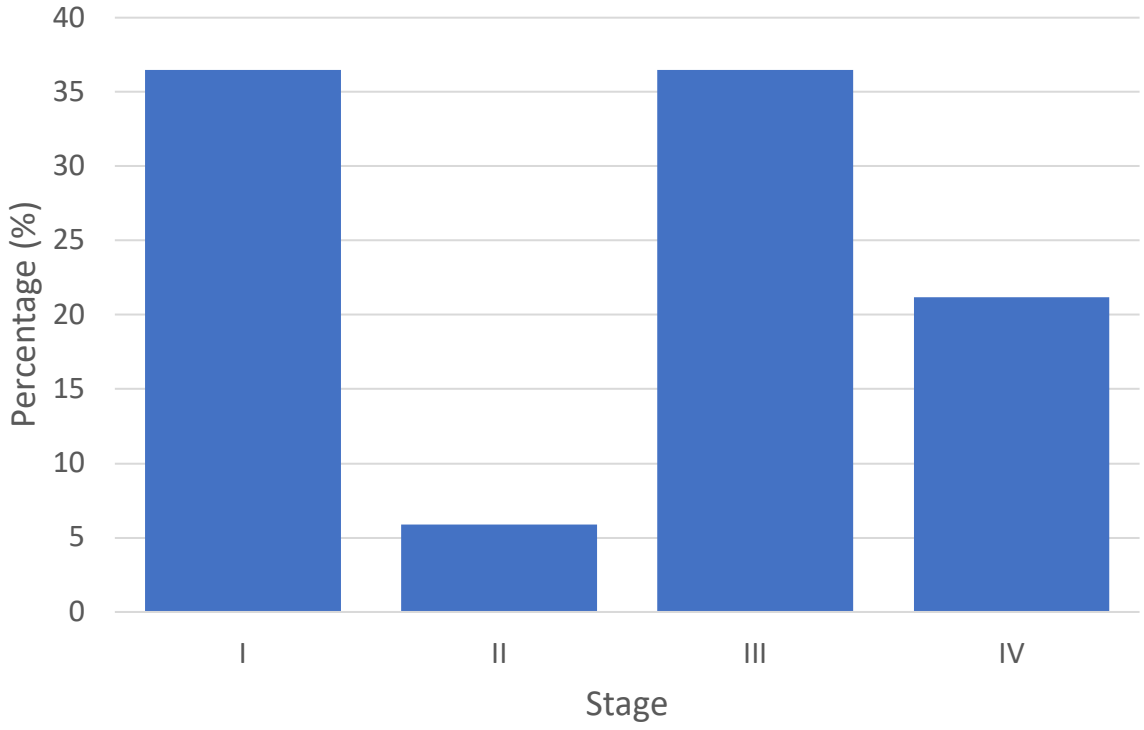


Figure 1.2. Stage distribution in women diagnosed in England (2010-2014) with a known stage (85%).[8]

1.3.3 Mortality and survival

In 2020, an estimated 207,252 women died from ovarian cancer worldwide.[1] Within the UK, the disease is the 6th most common cause of cancer death in women, accounting for over 4,000 deaths annually.[12] The prognosis for women diagnosed with ovarian cancer has improved markedly in recent decades. Women diagnosed with ovarian cancer in 2000-2001 in England and Wales had an estimated age standardised 5-year net survival of 38%.[13] By 2010-2011 this had increased to 46%. However, recent evidence from the ICBP indicates that UK age standardised net survival (3-year) is lower than four (Australia, Canada, Denmark, Norway) of the six other high-income countries included in the ICBP comparative analyses.[8] The reasons for the variation in survival between the countries remain unclear. However, differences in patient healthcare beliefs, health seeking behaviours, access to diagnostic investigations, willingness of GPs to perform cancer investigations, national diagnostic pathways and approaches to, and availability of, cancer treatment, have all been suggested as possible explanations.[3,9,14,15]

Ovarian cancer survival falls dramatically with increasing stage at diagnosis, with estimated UK 5-year net survivals in England of 93% for stage I, 68% for stage II, 27% for stage III and 13% for stage IV disease.[16]

1.3.4 Risk and protective factors

In addition to age and ethnicity, a number of genetic and environmental factors influence ovarian cancer risk.

Inherited genetic susceptibilities are responsible for an estimated 5-15% of ovarian malignancies, with germline mutations in the breast cancer genes (BRCA1 and BRCA2) accounting for the majority of these cancers.[17–19] The average cumulative risk of ovarian cancer by age 70 is estimated at 39-59% in BRCA1 mutation carriers and 11-17% in BRCA2 mutation carriers.[20,21] By contrast, the estimated lifetime risk of ovarian cancer in UK women (born after 1960) has been estimated at 2%.[22] Women with Lynch syndrome, an autosomal dominant cancer predisposition syndrome resulting from germline mutations in DNA mismatch repair genes, are at increased risk of ovarian cancer in addition to a range of other malignancies including colorectal and endometrial cancer.[23] A retrospective cohort study estimated that the lifetime risk of ovarian cancer in women with Lynch associated mutations was 7%.[24] Risk of ovarian cancer is greater in women with family history of breast or ovarian cancers, with risk increasing if multiple first degree relatives are affected at a young age or at multiple sites.[25–27] In addition to germline BRCA and Lynch associated mutations, a wide variety of other genetic mutations and shared environmental factors also contribute to increased familial risk.[26]

Hormonal and reproductive factors affect ovarian cancer risk. Factors which contribute to a greater number of lifetime ovulatory cycles are associated with an increased risk of ovarian cancer while those which result in episodes of anovulation are associated with a decreased risk. For example, epidemiological studies indicate that early menarche and late menopause are associated with an increased risk of ovarian cancer,[25,28,29] whereas higher parity and use of combined oral contraceptives are associated with a decreased risk of ovarian cancer.[30–32] A meta-analysis of 45 epidemiological studies found that the relative risk in women taking

combined oral contraceptive medication for 15 years or more (versus never users) was 0.42 (95% confidence intervals [CI]: 0.36–0.49).[30]

In contrast to oral contraceptive use, use of hormone replacement therapy is associated with an increased risk of ovarian cancer. A meta-analysis, which included individual datasets from 52 studies, reported that the relative risk for current users (versus never users) was 1.43 (95% CI 1.31–1.56).[33] The study also found that risk decreased with length of time from treatment cessation, but was still elevated for specific types of ovarian cancer at 10 years post-cessation.

Epidemiological studies have reported positive associations between a variety of other factors and ovarian cancer risk, including greater body fatness,[34–37] greater adult attained height,[36–38] and asbestos exposure.[39] Smoking is associated with an increased risk of some specific types of ovarian cancer.[40] A link between talc exposure and ovarian cancer is controversial. The International Agency for Research on Cancer (IARC), which is part of the World Health Organisation (WHO), currently lists perineal talc use as a possible carcinogen (within the 'limited evidence' category).[41] However, a recent pooled analysis, which included over a quarter of a million women, did not identify a significant association between genital talc exposure with ovarian cancer risk.[42]

1.4 Tumour types

Rather than a single disease, the term *ovarian cancer* encompasses a heterogeneous group of diseases that occur in the same anatomical region but differ in their cells of origin, clinical behaviour, treatment and prognosis. Understanding of the pathogenesis of ovarian cancer has improved dramatically over the last two decades, which has led to changes in how the disease is defined. In this section, I provide an overview of the different types of ovarian cancer and give the operational definition of ovarian cancer I use throughout this thesis.

1.4.1 Histological classification

Ovarian cancers can be broadly classified, on the basis of tissue of origin, into three principal categories: epithelial, germ cell and sex-cord stromal (**Figure 1.3**). Other tumour types exist, such as ovarian sarcoma, but as these are extremely rare I do not discuss them further in this thesis.

The proportion of ovarian cancers which fall into each category varies internationally, but the epithelial tumour type is consistently responsible for the vast majority of ovarian cancers.[43] In England, epithelial cancers account for more than 90% of all ovarian malignancies and are responsible for the majority of ovarian cancer-associated mortality.[44] They can be classified into four principal categories based on cell type (**Figure 1.3**). The serous histo-type is sub-divided on the basis of tumour grade into low grade serous (LGS) and high grade serous (HGS). HGS is the most common type of ovarian cancer, accounting for around 70% of epithelial ovarian malignancies.[45]

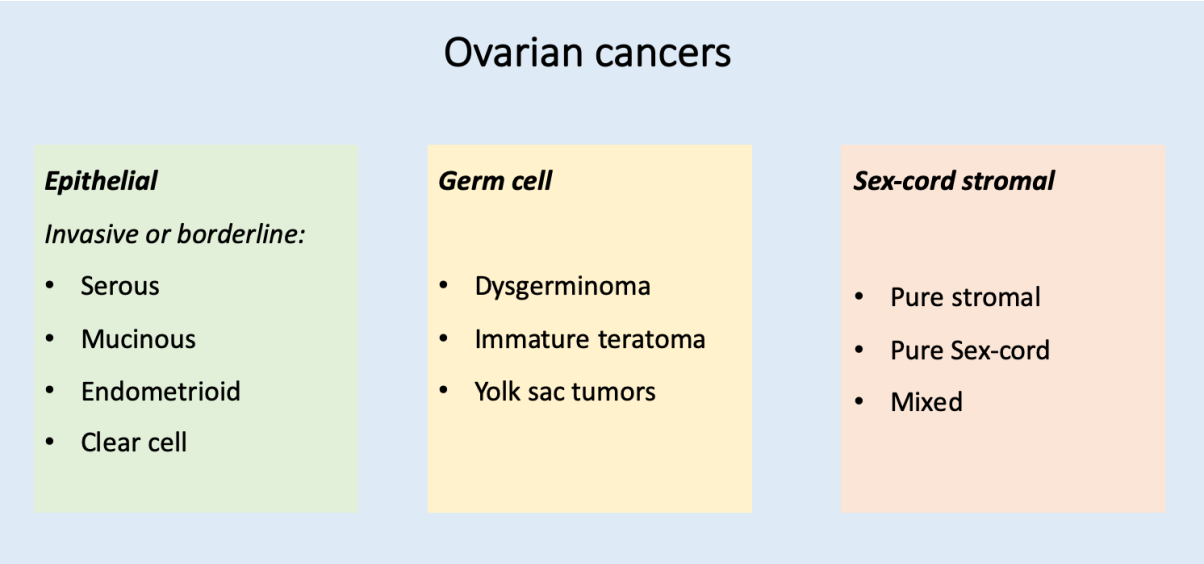


Figure 1.3. Ovarian cancer tumour types.

Malignant sex-cord stromal and germ cell tumours collectively account for under 5% of ovarian cancers diagnosed in England.[46] While epithelial malignancies usually occur in women over the age of fifty years, sex-cord stromal and germ cell cancers occur most commonly in younger women (<35 years of age), frequently affecting women in their late teens and early twenties.[46] The prognosis for women with these rare types of tumour is much better than for those with epithelial cancers. Even if diagnosed at an advanced stage, more than 75% of women with germ cell cancer are cured.[47]

1.4.2 Borderline tumours

Borderline is a type of tumour virtually unknown outside of the ovary. Borderline tumours are epithelial in origin and exhibit morphology and behaviour which lies between that of benign tumours (which cannot invade surrounding tissue and metastasise to other parts of the body) and overtly malignant/invasive tumours (which can invade surrounding tissue and metastasise to other parts of the body). In fact, when they were originally described in 1929, they were termed "semi-malignant" tumours.[48] Although borderline tumours do not penetrate and destroy surrounding tissue, they can exhibit weak invasive behaviour (termed micro-invasion).[49] At diagnosis, most are localised to the ovary (82%) with the remainder having spread to other sites within the abdomen and pelvis (distant metastases are rare).[49] Borderline tumours are treated surgically, but can recur (<10%), and approximately 5% transform into low grade serous cancer.[49,50] They principally affect women under the age of 45 years of age and account for 14% of all ovarian malignancies in England.[46]

Until the 3rd edition of the WHO International Classification of Diseases for Oncology (ICD-O3) was published in 2013, borderline tumours were included alongside other malignant tumours (which I hereafter refer to as invasive tumours for clarity) under the heading of "Malignant neoplasm of the ovary" (C56).[51] They are now categorised as "neoplasm of uncertain behaviour" (D39.1). Borderline tumours are staged in the same way using the same staging system as invasive tumours.[52]

Published early detection studies have varied as to whether they have included borderline tumours in their ovarian cancer outcome definition. Some, including the large UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), have included borderline tumours.[53,54] However, as survival for women with borderline tumours is good, even in late stage disease, many other studies have excluded borderline tumours and focussed only on invasive cancers.[3,53,55]

Ultimately, histological specimens are required to distinguish early invasive tumours from borderline tumours and both types of tumour are managed surgically.[56] Women presenting symptomatically to their GP due to the presence of either a borderline tumour or an invasive tumour should be referred urgently for specialist assessment and treatment.[57] Given this, I have included borderline tumours in my definition of ovarian cancer within this thesis.

1.4.3 '*Ovarian*' cancer: a misnomer?

All types of tumour discussed above were once thought to originate from the ovary and cancers of the fallopian tube were considered extremely rare. However, research conducted over the last two decades has provided compelling evidence that most (if not all) HGS epithelial ovarian cancers arise from the distal lining of fallopian tubes rather than from the lining of the ovary.[55,58–63] Some of the earliest evidence came from studies conducted in women undergoing risk reducing surgery (bilateral salpingo-oophorectomy) due to germline mutations in the BRCA genes. In these women, occult epithelial cancers were much more common in the fallopian tubes than in the ovaries in pathological specimens.[60,61] Non-invasive lesions, known as serous tubal intraepithelial carcinomas (STICs), were also identified in the fallopian tubes and genetic studies have since confirmed that STIC lesions are a precursor of HGS ovarian cancer.[62] Studies of pathological surgical specimens have also identified STIC lesions in around 60% of women with sporadic (i.e. non-germline BRCA associated) HGS ovarian cancers.[63]

A hallmark of ovarian and fallopian epithelial cancers is early spread to the peritoneum. Occasionally, epithelial cancers (most commonly HGS) are found to affect the peritoneum but not the ovary or fallopian tube. These are classified as primary peritoneal cancer.[51]

Regardless of where these cancers arise – ovary, fallopian tube or peritoneum – they behave and are managed clinically in the same way. A single classification system, published by the Fédération Internationale de Gynécologie et d'Obstétrique (FIGO), is now widely used to stage cancers arising from these three sites.[52] Although recording of the primary site is encouraged during staging, this is often not possible as the cancer usually envelops ovarian and fallopian tube structures or is disseminated at surgery. Cancers of ovarian, fallopian and peritoneal origin are now usually included together in clinical trials and epidemiological studies.[9,64] These cancers are often referred to collectively in the literature as ovarian cancer.[55,65] For simplicity, I have adopted this approach throughout my thesis.

1.4.4 Ovarian cancer: operational definition

Within this thesis I define ovarian cancer as:

A tumour of borderline or invasive behaviour arising from the ovary, fallopian tube or peritoneum.

This broad definition is in line with national guidelines on ovarian cancer detection and initial management in England.[65]

1.5 Ovarian cancer detection

1.5.1 Early detection

The diagnosis of cancers earlier in their disease course is generally seen as key to improving patient outcomes including survival.[66,67] It is an important area for policy makers, with the recently published National Health Service (NHS) England long term plan setting the goal of detecting 75% of cancers at an early stage (I and II) by 2028.[68] It is clear that women diagnosed with earlier stage ovarian cancer have better survival than those diagnosed with later stage disease, both in the short (1-year) and longer term (5 and 10-years).[69] A key prognostic factor in later stage ovarian cancer (III and IV) is whether all gross tumour can be removed at surgery (complete cytoreduction).[70] Even in patients with the same stage disease, complete cytoreduction is associated with improved survival.[70–72] Clinicians and researchers have suggested that the detection of less bulky ovarian cancer, which is more amenable to complete surgical resection, could help improve patient outcomes including survival even if a stage shift is not achieved.[55] A lower proportion of women with stage IV disease (38%) undergo surgery in the UK than women with stage II-III disease (71%).[73] This may be due to patient choice (after consideration of the low cure rate in advanced disease), or a clinical decision based on the low chance of complete cytoreduction or the poor performance status associated with advanced ovarian cancer. Even in advanced stage ovarian cancer, detection of less bulky and debilitating disease may allow more women to undergo treatment with curative intent.[74]

1.5.2 Timely detection

Given the association between earlier stage at diagnosis and higher survival rates, it would seem reasonable to assume that the shorter the time between a patient first presenting to their GP with symptoms and being diagnosed (the diagnostic interval) the better the outcomes. However, the relationship between diagnostic interval and cancer outcomes is

complex. For example, a retrospective analysis of the healthcare records of 601 Canadian women with epithelial ovarian cancer reported that the median diagnostic interval was 35 days longer in women with stage I-II disease than in women with stage III-IV disease.[75] This may be because cancer patients with advanced disease have more severe symptoms and are sicker at presentation and therefore experience an expedited diagnosis (sometimes called the 'sick quick' phenomenon).[76] A questionnaire based study (n=560), performed as part of the ICBP, found that the odds of being diagnosed with advanced ovarian cancer increased with longer primary care interval (time between symptomatic presentation in primary care and referral) up to around 50 days then decreased, although the relationship was not statistically significant.[77] A retrospective interview-based cohort study of 1,381 Australian women with ovarian cancer found no association between time from symptom onset to diagnosis and survival.[78] However, observational studies which examine associations between diagnostic interval and survival can be affected by the waiting time paradox (in which patients diagnosed quickly have poor outcomes).[79] A recent study by Round *et al* reported that, for the four most common cancers (ovarian cancer was not examined), there was an association between higher use of the urgent suspected cancer referral pathway by GP practices, early stage diagnosis (except for colorectal cancer) and lower mortality.[80] This provides some support for the assumption that prompt investigation and treatment of cancer patients could improve outcomes.

By contrast, there is little direct evidence that more timely diagnosis has an effect on outcomes in symptomatic ovarian cancer. The Diagnosing Ovarian cancer Early (DOvE) pilot study in Canada found that women diagnosed using a proactive symptom triggered testing initiative (in which women with relevant symptoms are encouraged to self-refer for a series of tests) had lower volume and more resectable disease than those diagnosed via the usual care route.[55] However, numbers were small (11 ovarian cancers were diagnosed via the proactive screening route). The definitive study results are still awaited.

Prompt diagnosis can have psychosocial patient benefits. A Danish retrospective questionnaire study, which included 188 women with ovarian cancer, found that longer time between symptom onset and diagnosis was associated with lower patient satisfaction and some measures of quality of life.[81]

A recent retrospective study, which utilised routinely collected primary care data, reported that the median diagnostic interval in the UK (excluding Scotland) was 56 days.[82] A quarter of these women had a diagnostic interval of 133 days or more.[82] Data from the 2010 National Cancer Patient Experience Survey in England indicates that a third of women had three or more pre-referral primary care consultations relating to possible symptoms of ovarian cancer prior to ovarian cancer diagnosis.[83] In the English National Cancer Diagnosis Audit, of women diagnosed with ovarian cancer in 2014, GP practices reported that some kind of avoidable delay in diagnosis (due to patient, clinician or system factors) had occurred in just under a third of women with ovarian cancer.[84] This was the third highest proportion for any of the 20 cancer sites examined. While ovarian cancer is a challenging disease to detect (as discussed in subsequent sections), this indicates that there may be opportunities to expedite diagnosis and that this may be beneficial.

1.5.3 Approaches: screening *vs* symptomatic detection

A significant effort has been made over the last 30 years to develop approaches to detect ovarian cancer as early as possible. Large randomised controlled screening trials of asymptomatic women have been performed in Japan, the United States and the UK.[53,85,86] To date, none of these have demonstrated a long-term survival benefit. However, the final long-term follow-up results from the largest of these trials (UKCTOCS) are awaited and are expected to be published later in 2021.

In the absence of screening programs, the majority of women are diagnosed following symptomatic presentation.[78,87] In countries such as the UK, where GPs play a gatekeeping role, the majority of patients are diagnosed following a symptomatic presentation in primary care.[87,88]

1.5.4 Symptomatic presentations

Case control studies, predominantly conducted in the United States and the UK, have demonstrated that a number of symptoms occur much more commonly in women with ovarian cancer (prior to diagnosis) than in controls.[89] These studies were ground-breaking as ovarian cancer had often been perceived as the 'silent-killer', an unhelpful label which is still used by the media today.[90]

In a case-control study by Hamilton *et al*, which utilised GP records from England, odds ratios for relevant symptoms were similar for those subsequently diagnosed with early (stage I and II) and late (III and IV) disease.[91] In a case control study by Goff *et al,* performed in the United States using patient survey data, symptoms were reported with equal frequency in those with early and late stage disease, although numbers with early stage cancers were small.[92] A retrospective cohort study in the United States, in which 622 women with ovarian cancer were interviewed about the symptoms they experienced prior to their diagnosis, found that women commonly reported experiencing symptoms in both early and late stage disease.[93] However, some symptoms were reported less commonly in early stage disease including abdominal pain (stage I-II: 47%, stage III-IV: 58%) and abdominal distension (stage I-II: 36%, stage III-IV: 46%). A recent multi-cancer English population based study, which utilised routinely collected healthcare data, found that in patients with abdominal pain (the most common ovarian cancer symptom) more than 60% were diagnosed with a stage I-III cancer.[94] Where multiple symptoms were present, stage I-III cancers were even more likely. Collectively, these studies indicate that some women with early stage disease do develop symptoms and present in clinical practice, which provides an opportunity for early ovarian cancer detection.

More than a dozen case control and cohort studies have investigated the association between symptoms and ovarian cancer diagnosis. While these studies are relatively consistent in reporting an increased risk of ovarian cancer diagnosis in women with abdominal distension, bloating and abdominal pain, there is considerable variation between studies as to which other symptoms are important.[89] For example, Hamilton *et al* (in their routinely collected data case-control study in England) found that appetite loss was significantly associated with a subsequent diagnosis of ovarian cancer, while Goff *et al* (in their questionnaire-based case-control study in the United States) found no significant association.[91,92] Such differences are likely to be due to variation in study populations and study design.[89]

The estimated predictive values of individual symptoms reported by studies have been modest. This is demonstrated by the estimated Positive Predictive Values (PPV) of symptoms identified by Hamilton *et al* (**Table 1.1**).[91] These symptoms are common in women presenting to their GP. Lim *et al* examined the GP records of almost 20,000 women (aged 45-74 years)

who visited GP practices in England and found that, over the course of a 12 month period, up to half reported a symptom of possible ovarian cancer.[95] So, while symptoms may occur more frequently in women with ovarian cancer, the vast majority of women with these symptoms will not have the disease.

Table 1.1. Symptoms associated with ovarian cancer in multivariable analysis from Hamilton *et al* (2009)[91]

Symptom	PPV (%)
Abdominal distension	2.5
Appetite loss	0.6
Postmenopausal bleeding	0.5
Abdominal pain	0.3
Abdominal bloating	0.3
Urinary frequency	0.2
Rectal bleeding	0.2

Determining which symptomatic women presenting to primary care should undergo specialist investigation for ovarian cancer is a diagnostic challenge. To help guide this decision, many countries (including the UK) have developed recommendations on the initial assessment and testing of symptomatic women for ovarian cancer, which I will discuss in the next section.

1.6 Initial assessment for ovarian cancer

At the outset of my doctoral research, I undertook a review of published clinical guidelines covering the initial assessment of women for ovarian cancer. By systematically searching relevant databases, I identified 18 guidelines which provided recommendations on the initial assessment and investigation of possible ovarian cancer in symptomatic women.[57,96–112] I compared these guidelines in terms of the symptoms they included, the physical examinations they recommended and the tests they advocated. I undertook this review to gain insight into current international recommended practice for the assessment of women for possible ovarian cancer in order to inform and contextualise my subsequent doctoral research studies.

A paper, based on this work, has been published:

Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. Garth Funston, Marije Van Melle, Marie-Louise Ladegaard Baun, Henry Jensen, Charles Helsper, Jon Emery, Emma J. Crosbie, Matthew Thompson, Willie Hamilton & Fiona M. Walter. BMC Cancer. 2019; 19:028

In the remainder of this section, I discuss the key results of this review in order to provide an overview of the international clinical context of this thesis. The review methods are summarised in **Box 1.2** and the related publication is included in **Appendix A** for reference. All tables and figures relating to my research within this section and subsequent sections of this thesis were prepared by me - no journal copy-edited versions are presented.

Search strategy

- *Searches*: Conducted between 1st-13th of March 2018 using key words and Medical Subject Heading (MeSH) terms relating to ovarian cancer and guidelines /consensus statements (any language)
- *Data sources*:
  - Medline, EMBASE and six guidelines databases
  - >20 professional / government websites hand searched

Selection criteria

- *Inclusion*:
  - Provided guidance on the initial assessment of women presenting with symptoms of possible ovarian cancer
  - Produced by a professional or governmental body
  - Published in the 10 years prior to the 13th March 2018
- *Exclusion*: Guidance solely covers…
  - management after a pelvic mass has been confirmed
  - screening
  - subgroups e.g. patients with BRCA mutations

Screening and selection

- *Titles and summaries*: Two reviewers independently screened titles and summaries*
- *Full text*: If potentially relevant at screening, two reviewers independently assessed the full text
- *Disagreements:* resolved by consensus

Data extraction

- Two reviewers independently extracted data against a specifically designed template*
- *Disagreements:* resolved by consensus
- *Data extracted included*:
  - Background details e.g. development body, country
  - Symptoms listed
  - Recommended examinations
  - Recommended tests

Data synthesis

- Data was summarised in tabular or graphical form and in prose

Box 1.2. Summary of review methods.

*English language documents were screened, and data extracted, by me and another reviewer (Dr Marije van Melle). For documents published in other languages, this was performed by pairs of study team members fluent in the relevant language.

1.6.1 Guidelines

Of the 18 guidance documents included in this review, two were developed in continental Europe (Germany and the Netherlands),[96,105] five in the UK and Republic of Ireland,[57,106–109] three in Scandinavia (Denmark, Norway and Sweden),[110–112] four in North America (United States and Canada)[101–104] and four in Australasia (Australia and New Zealand)[97–100] (**Figure 1.4**). Thirteen documents were published in English.

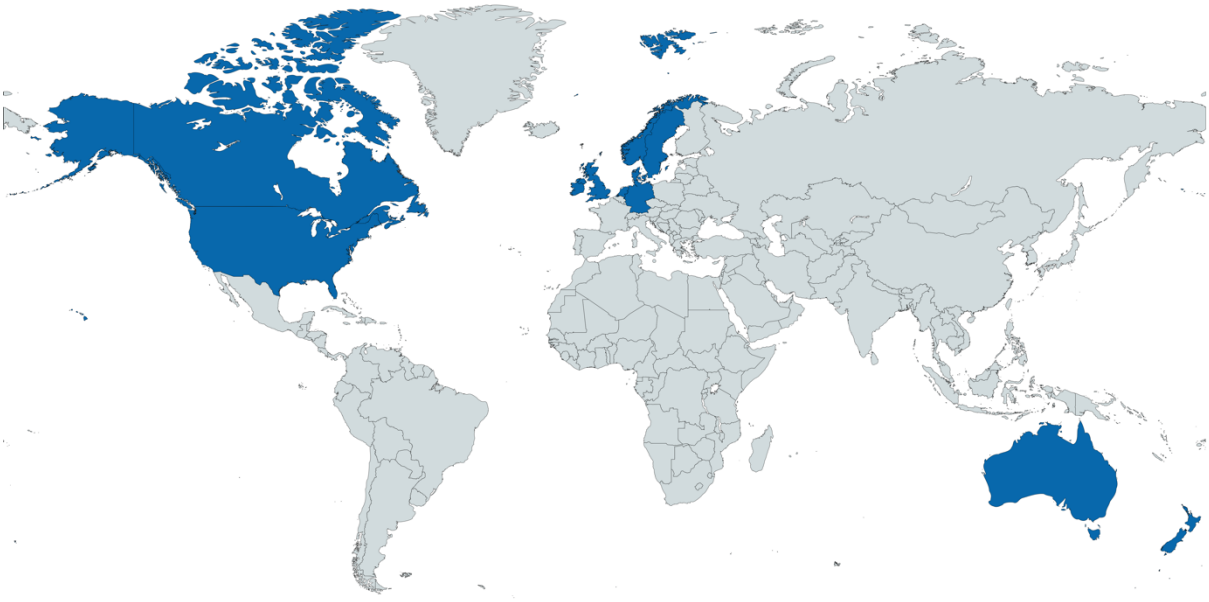


Figure 1.4. Countries of origin of included guidelines.
Figure created on mapchart.net under CC BY-SA licence.[113]

1.6.2 Symptoms and signs

All guidance documents provided advice regarding the symptoms that should prompt a healthcare professional to consider ovarian cancer (a review inclusion criterion). One or more of the related terms bloating, abdominal distention and increased abdominal size/girth, were listed in all guidance documents (**Figure 1.5**). Five further symptom terms were included in the majority of documents: abdominal or pelvic pain (n=16), urinary frequency (n=14), early satiety (n=14), change in bowel habit (n=12) and loss of appetite (n=10). The remaining 20 symptom terms were included in less than half of the documents. The number of ovarian cancer symptoms included in each document ranged from four to 14. Some documents simply listed symptoms which should prompt consideration for ovarian cancer while others detailed further factors which should be taken into account, including symptom frequency (e.g.

>12x/month), nature (e.g. persistent), duration (e.g. >1 year) and age at presentation (e.g. >50 years).

Fourteen documents recommended performing a physical examination and highlighted possible signs of ovarian cancer. Thirteen advocated abdominal examination or mentioned abdominal signs. Nine recommended pelvic or gynaecological examination, with three specifying that this should include a speculum examination and three a bimanual or digital examination. Three documents recommended performing a rectal examination. The most common examination findings mentioned were abdominal / pelvic mass (n=11) and ascites (n=8).

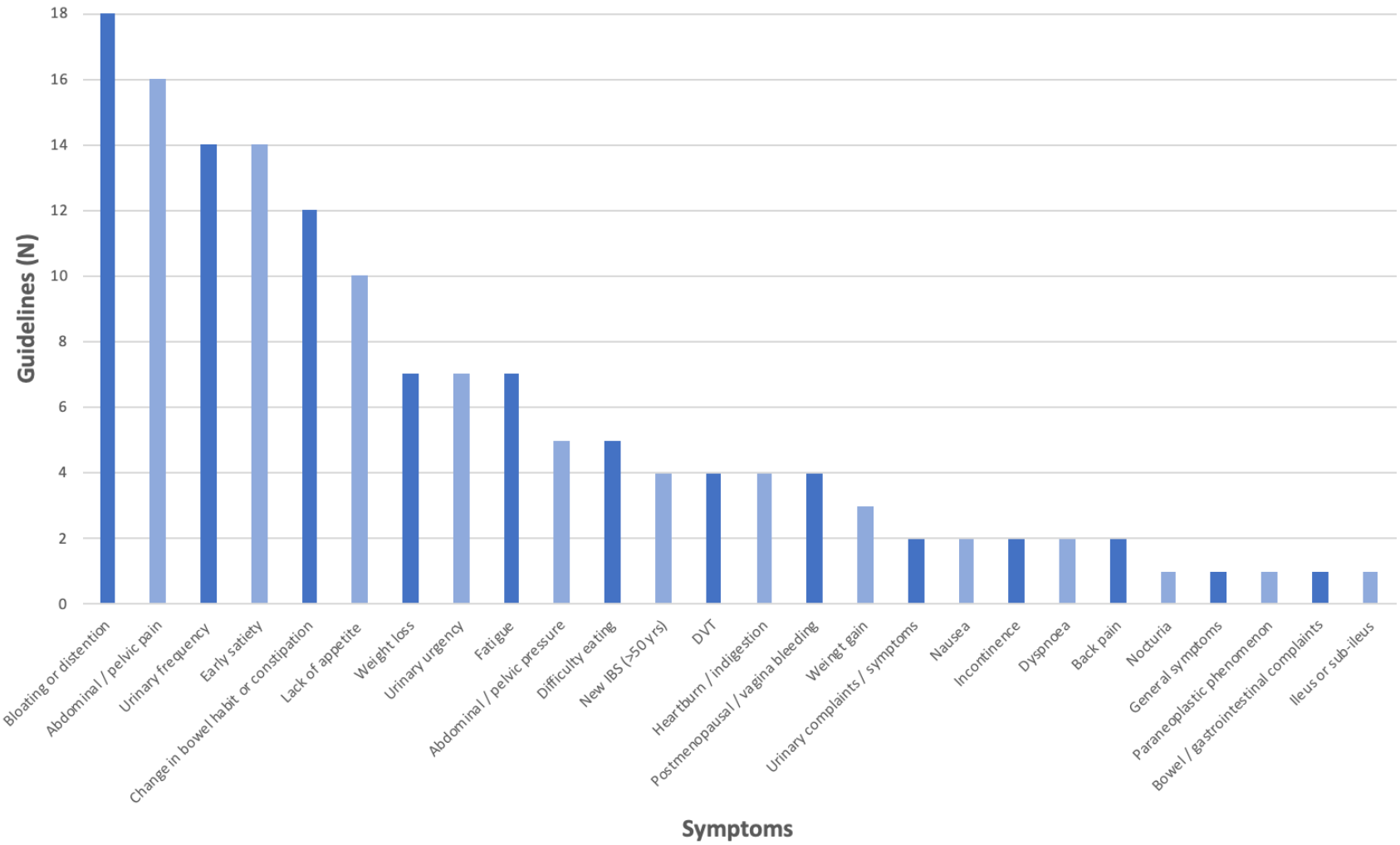


Figure 1.5. Frequency of inclusion of symptom terms in guidelines.

1.6.3 Tests

Guidance on the initial investigation of women with symptoms and or signs of possible ovarian cancer was provided in fifteen guidance documents. A wide range of testing strategies were recommended. I classified them (based on the number tests and the recommended order of testing) into one of five groups: 'single test' i.e. one test advocated; 'dual testing' i.e. performing two tests concurrently; 'sequential testing' i.e. performing a second type of investigation (second line) if the first type of investigation (first line) is abnormal; 'multiple testing options' i.e. where a range of investigation options were presented with no single investigation being advocated above another; and 'no testing' i.e. where no specific tests were recommended as part of the initial assessment. One document advocated a single test strategy; four a dual testing strategy; four a sequential testing strategy and three gave multiple testing options (**Figure 1.6**). Three did not recommend testing prior to specialist referral.

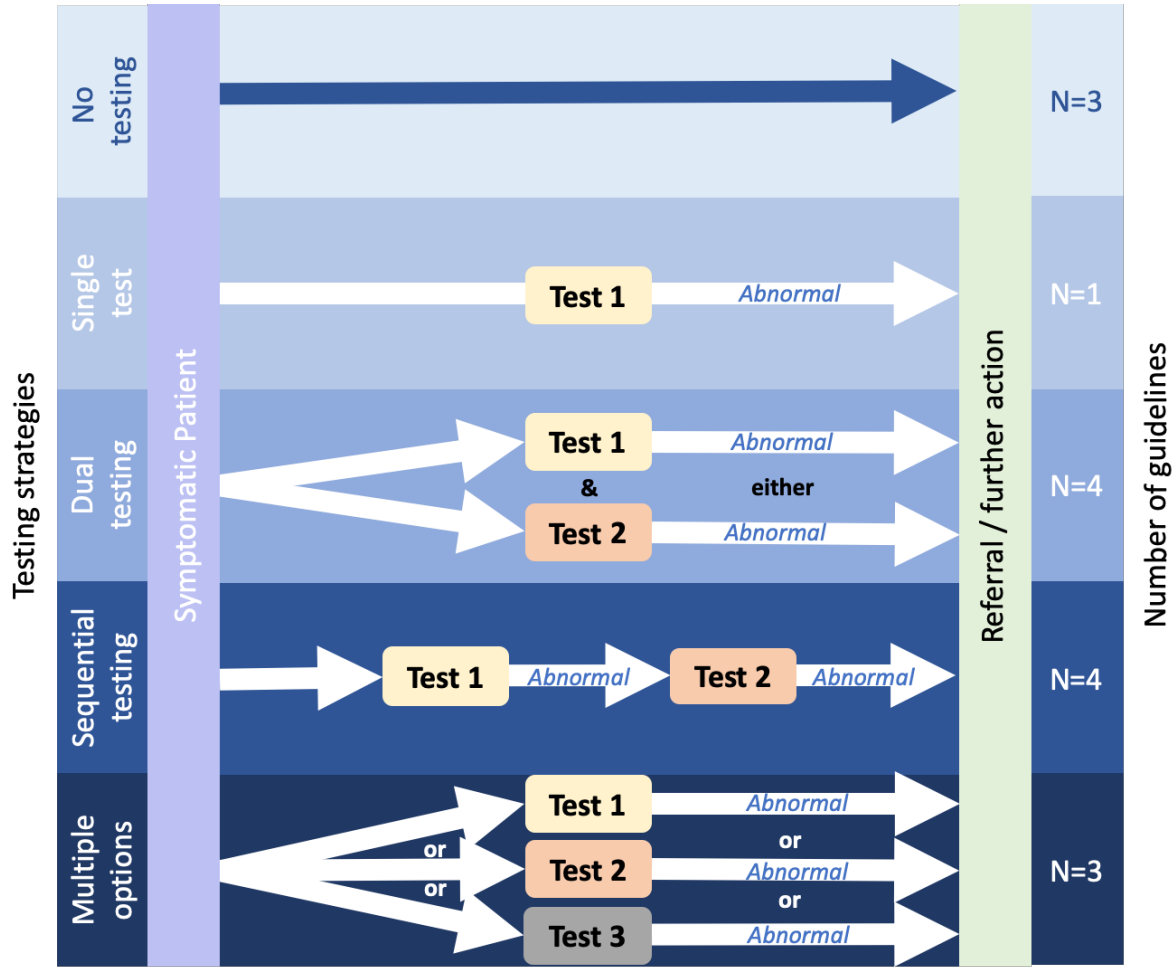


Figure 1.6. Recommended strategies for the investigation of women with symptoms of possible ovarian cancer.

The most commonly advocated tests for initial investigation were the serum ovarian cancer biomarker CA125 (11 documents) and ultrasound (12 documents) (**Table 1.2**). Four guidelines recommended that CA125 and ultrasound should be performed together (dual testing), three that CA125 should be performed first and an ultrasound only if the CA125 test is abnormal while one recommended the inverse – ultrasound then CA125 if an abnormality is found (sequential testing). Three simply listed the two tests as options for investigation.

Additional blood tests (e.g. full blood count) and imaging tests (e.g. Computed Tomography [CT]) were also recommended by several guidelines. Two diagnostic prediction models (based on CA125 level, ultrasound findings and age / menopausal status) were recommended by Australian guidelines: The Risk of Malignancy Index (RMI) and the Assessment of Different NEoplasias in the adneXa (ADNEX) model.

Table 1.2. Summary of tests recommended for the assessment of symptoms and/or signs of ovarian cancer.

Strategy	Guideline	When is testing advocated?	Initial tests
No testing prior to referral	Integrated ovarian cancer patient pathway (Denmark)	At point of specialist referral	***Note*** *CA125 requested in primary care at time of referral so as to be available to the specialist. Not acted upon in primary care*
	Ovarian cancer patient pathway (Norway)	Post specialist referral	Post referral
	Standardised ovarian cancer care pathway (Sweden)	At point of specialist referral	***Note*** *CA125 requested in primary care at time of referral so as to be available to the specialist. Not acted upon in primary care*
Single test	Guideline on diagnostics, therapy and follow-up of malignant ovarian tumours (Germany)	Signs or symptoms of OC	Transvaginal US ***Note:*** *CT, MRI, PET CT may be used in specific cases*
Dual testing	Scottish referral guidelines for suspected cancer (Scotland)	Symptoms of OC ***Note:*** *Ascites- refer urgently rather than test*	CA125 + pelvic US
	Management of epithelial ovarian cancer (Scotland)	Symptoms of OC	CA125 + pelvic US
	Assessment of symptoms that may be ovarian cancer: a guide for general practitioners (Australia)	Mass identified clinically ***Note:*** *No mass identified clinically- refer appropriately*	CA125 + transvaginal US ***Or*** CA125 + Abdominal US ***Or*** CA125 + CT
	Appropriate referral of women with suspected ovarian cancer (Australia)	Suspicious findings on clinical examination	CA125 + transvaginal US +/- calculation of Risk of Malignancy Index (RMI)

Sequential testing	Suspected cancer: recognition and referral (England)	OC symptoms ***Note:*** *Ascites or suspicious mass- refer urgently rather than test*	**First line:** CA125 **Second line:** Abdominopelvic US (if CA125 is abnormal)
	Epithelial ovarian / fallopian tube / primary peritoneal cancer guidelines: recommendations for practice (UK)	OC symptoms ***Note:*** *Pelvic or abdominal mass- refer urgently rather than test*	**First line:** CA125 **Second line:** Abdominopelvic US (if CA125 is abnormal)
	Ovarian cancer GP referral for symptomatic women (Ireland)	History suspicious of OC but examination normal ***Note:*** *Suspicious pelvis mass or ascites- refer urgently rather than test*	***First line*:** CA125 **Second line:** US of pelvis (If CA125 35-200 U/ml) ***Note:*** If CA125 >200 U/ml refer without US
	Ovarian cancer diagnosis pathway map (Ontario, Canada)	Suspicion of OC ***Note:*** *Tests may be performed prior to specialist referral but are not a requirement for referral. Can refer prior to testing*	**First line:** Transvaginal US and / or other imaging **Second line:** CA125, FBC, Renal Function + RMI *(**If indicated:** CEA, CA19-9, other tumour markers e.g. AFP, LDH, HCG)*
Multiple testing options	Optimal care pathway for women with ovarian cancer (Australia)	Symptoms of OC	Pelvic US + Routine blood tests + CA125 + Algorithms such as RMI, ADNEX +/- CT scan
	Genital tract cancers in females: ovarian, fallopian tube, and primary peritoneal cancers (British Columbia, Canada)	Suspicion of OC ***Note:*** *Imaging not essential for referral*	Transvaginal or abdominal US Blood tests: CA125**,** CA19-9**,** CA15-3, CEA ***<40 yrs old:*** AFP, HCG, LDH

	Ovarian cancer Including fallopian tube cancer and primary peritoneal cancer (USA)	Suspicion of OC ***Note:*** *Provides some advice on when particular tests are indicated. Appears to include both initial and pre-surgical tests*	US *and/or* abdominal/pelvic CT/MRI (as indicated) Chest CT or chest x-ray (as indicated) Complete blood count, chemistry profile and LFT CA125 or other tumour markers (as indicated: inhibin, β-hCG, AFP, LDH, CEA, CA19-9) Nutritional status GI evaluation (as indicated)
Unclear or no recommendations on testing given	Suspected cancer in primary care: guidelines for investigation, referral and reducing ethnic disparity (New Zealand)	No recommendations	No recommendations
	The role of the obstetrician-gynaecologist in the early detection of epithelial ovarian cancer in women at average risk (USA)	No recommendations	No recommendations
	Epithelial ovarian carcinoma (Netherlands)	Suspicion of OC. Not clear which tests should be used and when they should be used for initial investigation	Blood tests discussed: routine blood tests, CA125 +/- CEA

OC = ovarian cancer, US = ultrasound, MRI = magnetic resonance imaging, CT = computed tomography, PET = positron emission tomography, FBC = full blood count, CEA = carcinoembryonic antigen, CA19-9 = cancer antigen 19-9, AFP = alpha-fetoprotein, hCG = human chorionic gonadotropin, LDH = lactate dehydrogenase (LDH), GI = gastrointestinal.

1.6.4 Discussion of review findings

Some international variation in recommendations for the initial assessment and investigation of possible ovarian cancer is to be expected given the significant differences between healthcare systems. For example, the review included guidelines from countries with national health services and private healthcare systems and from two countries where direct access to gynaecology is available. However, the review does demonstrate that there is no clear international consensus on what clinical features should prompt clinicians to consider ovarian cancer, what physical examinations should be carried out and what initial test(s) should be performed.

The majority of guidelines included the four symptoms with the highest positive likelihood ratios for ovarian cancer in a recent systematic review by Ebell *et al* (abdominal distension, abdominal/pelvic pain, bloating and appetite loss).[89] However, a wide range of other symptoms appeared in guidelines including those which were not found to substantially increase the likelihood of ovarian cancer in Ebell *et al's* review e.g. nausea and back pain. Variation in the symptoms which appear in guidelines reflects the wide variation in the symptoms which studies have identified as being associated with ovarian cancer over the last two decades. Some guideline producers, such as NICE (the body which produces clinical guidelines for the NHS in England, Wales and Northern Ireland) and the Scottish Intelligence Guidelines Network (SIGN), followed a rigorous approach in their selection of which symptoms to include, informed by systematic reviews. Other guideline producers selected symptoms based on a single study or consensus approach.[§]

Although most guidelines recommended CA125 and ultrasound as tests for ovarian cancer, how they recommended using them varied considerably. Guidelines commonly mentioned the paucity of evidence on the diagnostic performance of CA125 and ultrasound for the detection of ovarian cancer in symptomatic women. For example, the NICE guidelines highlighted that they were not able to identify any studies comparing the performance of CA125 and ultrasound in primary care.[114] Other guidelines mentioned that due to lack of

[§] When the AGREEII tool was used to assess the quality of development and reporting of included guidelines, eight where identified as 'unsatisfactory' (a score of <50%) in the 'Rigour of Development' domain. This highlights potential issues with the quality of the development process for these guidelines (**Appendix A**).

evidence they used a consensus approach when making decisions about tests.[108,109] One did not recommend any particular test due to the lack of evidence.[100] How CA125 and ultrasound are used is likely to have consequences for ovarian cancer detection. A sequential testing approach (where both tests need to be abnormal to trigger specialist referral)[57] may be specific but at the cost of lower sensitivity. Conversely, a dual-testing approach (where an abnormality in either test warrants referral)[108,109] may be more sensitive but at the detriment of specificity.

1.6.5 Review summary

Recommended approaches for the initial assessment of ovarian cancer differ around the world. Two tests for ovarian cancer predominated in guidelines: ultrasound and CA125. However, recommendations differed on how and when these tests should be used. Lack of evidence on test performance in symptomatic women presenting in primary care was cited by several guideline producers as a key issue when determining which investigation(s) to include.

1.7 NICE guidelines in depth

In 2011, NICE published Clinical Guideline 122 (CG122): "Ovarian cancer: recognition and initial management".[65] The choice of symptoms and tests included was informed by a comprehensive review of the literature.[114] In 2015, when multi-cancer guidance was published by NICE (NICE Guideline 12 [NG12], "Suspected cancer: recognition and referral"), CG122 recommendations on ovarian cancer detection were incorporated unaltered.[57]

NICE recommendations on assessment, testing and referral for ovarian cancer are summarised in **Figure 1.7**. In women with symptoms which suggest ovarian cancer, NICE advocate a sequential testing approach in which CA125 forms the first-line test. If the CA125 is abnormal (≥35 U/ml), the guidelines recommend performing an ultrasound. If the post-CA125 ultrasound is abnormal, they recommend urgent referral to gynaecology via the suspected cancer pathway (a 'two week wait' referral). If CA125 is normal, or if CA125 is abnormal but ultrasound is normal, they recommend assessing for alternative symptom causes.

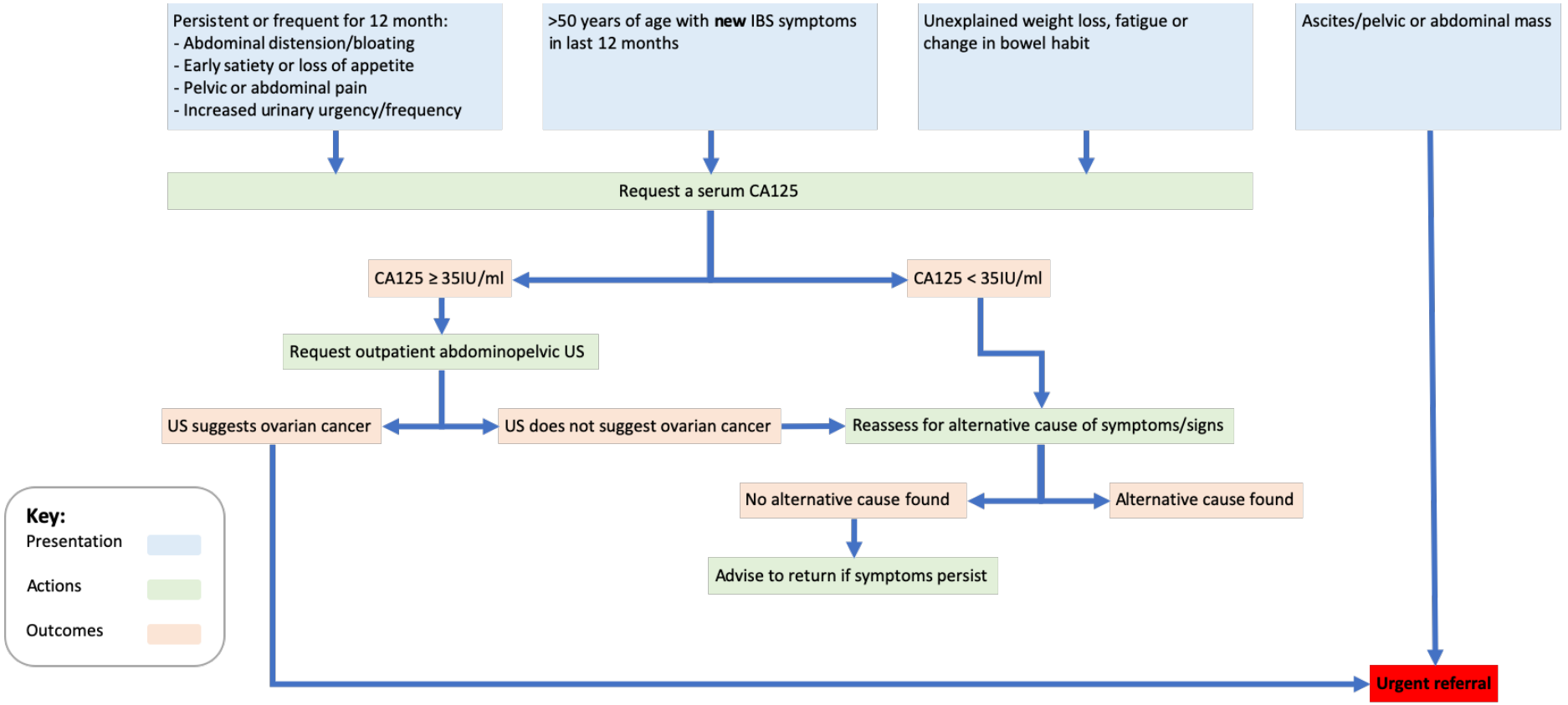


Figure 1.7. NG12 recommendations for the investigation and referral of suspected ovarian cancer.
Adapted from Funston *et al* (2018).[115]

1.8 Cancer Antigen 125 (CA125)

As described in the previous section, CA125 is recommend as an initial test for ovarian cancer in women with relevant symptoms in England and a number of other countries around the world.[116] In this section, I provide some background on the CA125 test, summarise current knowledge of its diagnostic accuracy and discuss how frequently it is used within English primary care.

1.8.1 What is CA125?

In 1981 Bast *et al* developed a monoclonal antibody (ovarian cancer antibody 125 [OC125]) which bound to an antigen on a high molecular weight glycoprotein (MUC-16) on the surface of ovarian cancer cells.[117] This antigen was later named CA125 (variously known as carbohydrate antigen 125, cancer antigen 125).[118] It was initially thought that CA125 was not expressed by normal epithelial ovarian tissue, but research has shown that the protein is present on the epithelial surface of healthy ovaries in addition to other healthy epithelial tissues including fallopian, endometrial, pancreatic, peritoneal, pleural, gastrointestinal and lung tissue.[119,120] Its biological role is still debated but it may be involved in lubricating and protecting epithelial surfaces.[121,122] In cancer, it is thought to play a role in metastasis and the protection of cancer cells from elements of the immune system.[121,122]

1.8.2 Serum CA125 levels

In a landmark study, published in the New England Journal of Medicine in 1983, Bast *et al* reported that CA125 could be detected in the blood and that levels were above a cut-off of 35 U/ml in 82% (83/101) of patients with ovarian cancer, 6% (9/143) of patients with benign diseases and 1% (9/888) of healthy women.[118] Based on this research, a CA125 cut-off of ≥35 U/ml was widely adopted and is still used today. Subsequent studies found that the proportion of women with CA125 levels ≥35 U/ml varies by ovarian cancer stage. A review, which included 723 women with ovarian cancer, found that CA125 was elevated in >90% of stage II-IV cancer but only around half of stage I ovarian cancer prior to surgery.[119] It is also less commonly elevated in women with mucinous histology, borderline tumours and non-epithelial tumours than in other tumour types

of ovarian cancer.[119,123,124] A large meta-analysis (N=2,374) found that, overall, CA125 levels are ≥35 U/ml in 80% of women with invasive or borderline ovarian tumours pre-surgery.[125]

CA125 is not specific for ovarian cancer. A range of benign gynaecological and non-gynaecological conditions can elevate CA125 levels (**Box 1.3**). Studies have also reported that CA125 can be elevated in a number of non-ovarian malignancies, notably pancreatic cancer (45-61%),[126–128] lung cancer (29-30%),[119,129] endometrial cancer (21-35%),[130,131] and colorectal cancer (10-24%).[132–134]

Benign Conditions
Endometriosis[135]
Fibroids[136]
Pelvic inflammatory disease[136,137]
Benign ovarian cysts[136,138]
Heart failure[139,140]
Chronic liver disease[141]
Non-ovarian malignancies
Endometrial cancer[130,131]
Pancreatic cancer[126–128]
Gastrointestinal cancer[119,132–134]
Lung cancer[119,129]
Breast cancer[119,142]

Box 1.3. Benign and malignant conditions associated with elevated CA125 levels.

In apparently healthy individuals, baseline CA125 levels are higher in premenopausal than postmenopausal women.[143] Levels can fluctuate during pregnancy and the menstrual cycle.[144,145]

1.8.3 Clinical roles for CA125

In addition to its role in detection of ovarian cancer in primary care, CA125 is widely used to monitor progress in women with ovarian cancer undergoing treatment following diagnosis and to detect relapse.[120] It is also used within algorithms, such as the RMI, to help determine whether an adnexal tumour is benign or malignant to guide treatment decisions, as women with likely ovarian cancer should be managed by a gynaecological oncologist in a tertiary centre.[65]

Studies have investigated potential roles for CA125 in diagnosis, prognosis and monitoring of a range of benign conditions. For example, it has been evaluated as a diagnostic and prognostic marker for endometriosis[135,146] and heart failure.[139,147] It has also been evaluated, in isolation and within biomarker panels, for the detection of cancers other than ovarian e.g. pancreatic cancer.[126–128,148] Despite this research, it is not routinely used as a test for any diseases other than ovarian cancer. This may change with the recent development of multi-cancer 'liquid biopsies'. For example, the CanerSEEK panel was developed as a single test for eight different cancers and combines 10 protein biomarkers alongside circulating tumour DNA. CA125 is included in CancerSEEK and was, in the original CancerSEEK case-control study, the most important protein biomarker in terms of CancerSEEK accuracy.[149] In a sensitivity analysis, removing CA125 from CancerSEEK reduced the sensitivity of the test not only for ovarian cancer but for liver cancer, oesophageal cancer and lung cancer.

1.8.4 The diagnostic accuracy of CA125

The diagnostic performance of CA125 has been evaluated both in secondary care patients with a known pelvic mass and in apparently healthy asymptomatic postmenopausal women within the screening context.

In women with an adnexal mass, a meta-analysis of 17 studies, which included a total of 2,374 women, reported that CA125 had a sensitivity of 80% and a specificity of 75% for ovarian cancer prior to surgery.[125] Studies have reported PPVs of 35-91% in this context.[150,151]

An early screening study, which included 1,010 postmenopausal women, found that performing a single CA125 test (cut-off ≥30 U/ml) had a specificity of 97%.[151] Results from the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial found that, when used annually in postmenopausal women, CA125 (≥35 U/ml) exhibited a sensitivity of 55%, and a PPV of 3.7%.[152] Subsequent screening studies have focused on the use of serial CA125 measurements (including within the Risk of Ovarian Cancer Algorithm [ROCA]) and of multimodal strategies combining CA125 testing with ultrasound.[53,153,154]

A comprehensive systematic literature review, performed to inform the development of NICE CG122 in 2011, did not identify any studies evaluating CA125 in primary care.[114] Subsequent surveillance reviews undertaken by NICE (most recently in 2020) have also failed to identify any relevant primary care evidence on CA125 diagnostic performance.[155] In order to identify any additional relevant papers which might provide insight on CA125 performance in primary care, I searched PubMed for studies published between database inception and the 1st July 2019, using the keyword "CA125" or its variants and "primary care" or its variants. I applied no language restrictions. This identified one study (not mentioned in the NICE surveillance review) which described outcomes in 152 women with CA125 levels ≥35 U/ml in primary care within the catchment population of a single UK hospital. This study reported that 16 (11%) women were diagnosed with ovarian cancer and a further 16 with another form of cancer.[156] I did not identify any other studies which provided insight into the diagnostic performance of CA125 in primary care.

1.8.5 Spectrum effect

The diagnostic performance of a test is influenced by the characteristics of the population in which it is being used. For example, secondary care studies have evaluated the performance of CA125 in women who all have either ovarian cancer, a benign pelvic mass or another gynaecological disease requiring surgery. These benign conditions are less prevalent in the primary care population. Where they are present in the primary care population they may, on average, be less severe. Benign ovarian masses and gynaecological conditions are known to be

associated with elevated CA125 levels. Therefore, it is likely that a greater proportion of women without ovarian cancer will have elevated CA125 levels in the specialist setting than the primary care setting due to the higher prevalence (and possibly greater severity) of benign CA125 elevating conditions. In women with ovarian cancer, the severity of disease (tumour burden) is likely to be greater in women in the specialist setting who are about to undergo surgery than in women in primary care at the start of the diagnostic pathway. This may result in fewer false negative results in the specialist setting than the primary care setting, as disease burden is associated with higher CA125 levels. Tests tend to have higher specificity and lower sensitivity when moved from a population with a high prevalence of the disease of interest to one with a low prevalence of the disease of interest: the spectrum effect.[157]

Similarly, the PPV of a test varies depending on the prevalence of disease within a population. It is lower in populations with lower prevalence of the disease of interest and higher in populations with higher prevalence. The difference in the PPV of a given test when it is used in high and low risk populations can be dramatic.[158] This can be crudely illustrated by comparing the PPV of a raised CA125 in women with a pelvic mass pre-surgery (35-91%),[150,151] to asymptomatic women in the PLCO screening trial (3.7%).[152] An understanding of the PPV and Negative Predictive Value (NPV) of a test within a given population is crucial to allow clinical interpretation i.e. what a normal and abnormal test mean in terms of a patient's probability of having a disease. Without knowing the PPV of CA125 in primary care, patients and clinicians cannot make fully informed decisions about the need for further testing or referral. In addition, NICE uses PPV when making recommendations about cancer testing and referral in symptomatic patients. When developing NG12 in 2015, they recommended urgent cancer referral in symptomatic primary care patients where the PPV for an individual cancer was ≥3%.[57] Prior to this thesis, no large studies had reported the PPV of a positive CA125 result in symptomatic women in primary care.

1.8.6 How common is primary care CA125 testing?

There is limited data on the number of women who have a CA125 test performed in the UK each year. A laboratory-based study estimated that in Oxfordshire, over a 3-month period at the start

of 2013, 70 / 10,000 women aged 50 years or older had a first CA125 test performed in primary care.[159] The rate was lower for women aged 25-49 years (50 / 10,000 women).

Not all women diagnosed with ovarian cancer have had a CA125 test performed in primary care prior to diagnosis. Using routinely collected primary care data from the Clinical Practice Research Datalink (CPRD) GOLD dataset, I found that, out of 1,223 women diagnosed in England (between 2012 - 2015), 582 (48%) had a CA125 test performed in primary care in the year before their diagnosis.[160] Some of the women without a CA125 test will have been diagnosed after presenting to the emergency department or incidentally in secondary care.[88] Some may also have been referred via the urgent suspected cancer pathway without a CA125 following identification of ascites or a mass in primary care (in line with NICE guidelines). In addition, GPs do not always adhere to guidelines,[161] and some GPs may have performed an ultrasound (without a prior CA125 test) and referred after an abnormality was detected on the scan.

1.9 Models and tools

A multivariable prediction model is defined as a "*mathematical equation that relates multiple predictors for a particular individual to the probability of or risk for the presence (diagnosis) or future occurrence (prognosis) of a particular outcome*".[162] Such models can be used to guide the clinical management of patients. For example, the QRisk models predict the risk of a patient developing cardiovascular disease within 10 years.[163] NICE recommends offering statin treatment if the QRisk model indicates that a patient's 10-year risk of developing cardiovascular disease is ≥10%.[164] These models can be integrated into GP IT software in the form of electronic clinical decision support (eCDS) tools which alert the clinician when a given risk level is reached. Examples of currently available primary care cancer eCDS tools include QCancer and the electronic Risk Assessment Tool (eRAT) for cancer.[165,166] These tools incorporate symptom information alongside patient demographics / risk factors (e.g. age) and test results (e.g. haemoglobin level and platelet count) in order to identify patients at a higher risk of undiagnosed cancer. The impact of these tools on patient outcomes (including diagnostic interval, stage at diagnosis and survival) remains unclear,[167] but a pragmatic cluster randomised control trial is

currently underway to evaluate the effect of eRAT implementation on stage at cancer diagnosis and survival in multiple cancers (including ovarian).[166]

In the screening and secondary care settings, a variety of multivariable models incorporating CA125 have been developed. It is these models, rather than the CA125 level alone, that are generally used to guide patient care. For example, the RMI (which includes menopausal status, CA125 level and specific ultrasound findings) is used in secondary care within the UK to decide whether patients should be managed in a tertiary care setting by a gynaecological oncologist.[65] Other secondary care models incorporating novel ovarian cancer blood tests, such as Human Epididymis protein 4 (HE4),[168] have been developed for the same purpose. These models are not appropriate for UK primary care, as they require specialist interpretation of ultrasound data and / or specialist blood test results.

In the prostate cancer screening setting, multivariable models which incorporate patient demographics, examination findings and other blood tests alongside Prostate Specific Antigen (PSA), have been shown to outperform PSA alone in the detection of cancer.[169] It is possible that that similar models incorporating CA125 along with other information available at the point of CA125 testing in primary care, including patient demographics (e.g. age), presenting symptoms and the results of routine blood tests, could outperform CA125 alone in the detection of ovarian cancer.

1.10 Thesis methods

I considered using different methodological approaches in order to achieve the thesis aim of evaluating the diagnostic performance of CA125 in primary care and developing and evaluating novel approaches to improve its performance. In this section I discuss the methodological options and my rationale for employing a cohort design using routinely collected data for all empirical studies within this thesis.

Diagnostic accuracy studies frequently employ a prospective study design, in which appropriate participants are recruited by the research team to undergo the index test which is then compared

against a gold standard reference.[170] A similar approach can be used in diagnostic modelling studies where model variable data is prospectively collected by researchers.[171] Such an approach has the advantage of allowing participant selection and data collection to be carefully tailored to the study question. However, prospective data collection can be time consuming, expensive and logistically challenging, particularly when the outcome is rare.[171] For example, in one study which sought to evaluate a symptom triggered testing approach for ovarian cancer detection, 5000 women were approached in primary care clinics and screened for symptoms of possible ovarian cancer, with further investigations performed (CA125 and ultrasound) if relevant symptoms were present.[172] However, only two ovarian cancers were identified in the cohort so few conclusions could be drawn.

An alternative approach to prospective data collection is to use existing routinely collected data sources. In England, a number of routinely collected primary care data sources, which contain coded clinical information including test results, are available for use in research. In addition, information on cancers diagnosed in England is systematically collected at a national level and collated within a cancer registry, enabling researchers to identify which patients within their study have cancer.[173] Such data sources have previously been used both to evaluate tests which, like CA125, are already in routine clinical use and to develop and validate novel cancer diagnostic prediction models.[54,174–177] A major advantage in using these data sources is their large size: several UK primary care data sources each contain records for over 10 million patients,[178–180] which has enabled the study of outcomes, such as ovarian cancer, that are relatively rare in primary care.[54,181] Where large numbers of study participants or large amounts of data are required, use of routinely collected data may be more economical (both financially and in terms of time) than prospective data collection.[171] Routinely collected data studies may also better reflect the performance of a test as it is used in standard clinical practice than prospective studies, in which selection criteria are often rigorously applied by the research team.[170]

While using routinely collected data to evaluate tests and to develop prediction models does have limitations, which I consider in this thesis, taking account of resource constraints and the

relative rarity of ovarian cancer in primary care, I judged that it was the most appropriate approach to achieve my thesis aim. I describe the data sources used in this thesis in depth in **Chapter 2**. In the appropriate chapters and sections, I discuss limitations arising from the use of routinely collected data including issues with data coding and missing data (**Sections 2.5, 2.8, 3.3.2, 3.5.2, 4.4.1, 6.2, 6.4.2, 7.2**) participant selection (**Sections 3.3.2, 6.2.5**) and determination of clinical outcomes (**2.5, 2.6, 3.4.1**).

Both cohort and case control study designs have previously been used to evaluate the diagnostic performance of tests. However, case-control studies can overestimate test diagnostic accuracy,[182,183] and methodological guidelines discourage their use in diagnostic accuracy studies for this reason.[170,184] In addition, while the PPV and NPV of a test for a given population can be estimated using a representative cohort of patients identified from that population, they cannot reliably be determined using a case control design, as the prevalence of the outcome may not reflect the prevalence of the disease within the population of interest. Given these considerations, I adopted a cohort study design in this thesis. The specification of this cohort is discussed in detail in **Chapter 2**.

1.11 Summary

Ovarian cancer is a relatively rare diagnosis in primary care, but it is the 6th most common cancer and the 6th most common cause of cancer death in UK women.[2,12] There is evidence to indicate that expediting diagnosis in symptomatic women could improve patient outcomes.

CA125 is recommended as the single first line test for ovarian cancer in women with possible symptoms of the disease presenting in primary care in England.[57] The test is also recommended as part of the initial testing strategy in other countries including Scotland, the Republic of Ireland, Australia, Canada and the United States.[116] However, the diagnostic accuracy of CA125 for the detection of ovarian cancer in primary care has not previously been established. The current CA125 cut-off of 35 U/ml (which is used in primary care internationally) was not chosen based on primary care evidence. Given the differences in populations, the diagnostic accuracy of CA125 in

primary care cannot be inferred from studies conducted within secondary care or screening settings. In order to understand how well CA125 performs in primary care it must be studied within a primary care population. In their 2020 surveillance report (and in previous surveillance reports) NICE called for research evaluating the diagnostic performance of CA125 in primary care in order to inform future guideline development.[155]

In the secondary care setting, diagnostic prediction models have been developed which include patient demographics, imaging findings and novel biomarkers, alongside CA125. Models incorporating another common cancer biomarker, PSA, with information including patient demographics and examination findings, outperform PSA alone in the screening setting. It is possible that a prediction model incorporating CA125, alongside other variables available within primary care at the point of CA125 testing, could outperform CA125 alone in the detection of ovarian cancer.

Studies have shown that CA125 can be elevated in a range of cancers other than ovarian cancer. However, the proportion of women with high CA125 levels in primary care due to these cancers has not previously been determined.

In the following chapters, I describe work undertaken to address gaps in our understanding of CA125 performance in primary care and approaches to improve its performance and clinical utility.

Chapter 2. Data sources and data preparation

2.1 Overview

A single dataset consisting of linked routinely collected data from CPRD GOLD, CPRD Small Area Levels, the National Cancer Registration and Analysis Service (NCRAS) and the Hospital Episodic Statistics Admitted Patient Care (HES APC) databases was obtained to conduct all primary research studies in this thesis (**Chapters 3, 4** and **6**). In this chapter, I first provide a brief overview of each of these data sources and describe the regulatory approval I obtained to use data from them for this thesis. I go on to define the baseline patient cohort. This directly formed the study cohort for **Chapter 3** and subsets of patients were selected from this cohort for the research presented in **Chapters 4** and **6**. I describe the steps taken to prepare this baseline cohort and key variables used across multiple chapters. Where a variable is used only in a single chapter, relevant definitions and data preparation steps are described in that chapter.

2.2 Data sources

In this section I provide a summary of the data sources used within this thesis.

2.2.1 The CPRD

CPRD GOLD (subsequently referred to simply as CPRD) contains longitudinal, anonymised, coded, primary care data for 11 million UK patients, of whom 4.4 million were 'active' (alive and registered with a participating GP practice) in July 2013.[178] It is broadly representative of UK demographics in terms of sex, ethnicity and age.[178] CPRD data has been used in over 2,600 research papers including many cancer diagnostic studies.[185]

Individual GP practices sign up to contribute data to the CPRD. Data on all patients from participating practices are then included unless a patient actively opts out. Data are uploaded from practices to the CPRD on a regular (usually monthly) basis. As well as practice level

information, including geographical region, CPRD contains patient level demographics and detailed clinical information. In the UK, most clinical information within GP records, such as symptoms, diagnoses, immunisations and tests, is coded. Until recently Read codes were used; although a new coding system (SNOWMED) is now being applied, its implementation postdates the data collected for this thesis. Read codes are a hierarchical clinical classification system.[186] Each Read code is associated with a Read term which allows interpretation e.g. Read code "44a6.00" is associated with Read term "CA125 level", and these map directly (1:1) to an internal CPRD coding system (Medcodes). Depending on the type of clinical activity, additional information can be recorded alongside the code within the CPRD. For example, in the case of blood tests, numerical test results and units of measurement may accompany the code. Each clinical code within the CPRD is dated and recorded against a unique patient identifier.

CPRD data are subject to quality assurance procedures at both the patient and practice level. Patients are flagged as either 'acceptable' or 'unacceptable' for research based on an assessment of their records including checks for invalid age or gender and invalid clinical dates e.g. clinical information recorded prior to their date of birth.[178] At a practice level, the quality of death recording and an assessment for gaps in recording is used to allocate an 'up-to-standard' (UTS) date, indicating when practice data is considered of sufficient quality for research.[178] Despite these checks, the CPRD comprises data that was collected for clinical use rather than for research and its quality largely depends on the completeness and accuracy of coding within GP practices. Clinicians do not always code clinical information such as symptoms or diagnoses, which may instead be recorded in free text within a record. Since 2013, this free text information has not been collected by the CPRD and is no longer available for research purposes.[187] However, the completeness and accuracy of blood test results within the CPRD is high as results are automatically transferred to the GP system from labs.[178]

CPRD data can be linked at an individual patient level to a range of external data sources by a third party - NHS Digital.[188] NHS Digital use a multi-step approach to match patients between data sources based on key variables (NHS number, date of birth, sex and postcode).[188] Linkage to

external data sources can improve data completeness and allow for the analysis of information which is not recorded within the CPRD. For example, studies can obtain additional information on hospital admissions by linking CPRD data to HES or on mortality by linking to the Office of National Statistics death registration dataset. However, linkage is only available for a subset of practices in England (~58% of all CPRD practices). For the purposes of this thesis, a CPRD dataset was linked to NCRAS, HES APC and Small Area Level data.

2.2.2 The NCRAS

The NCRAS acts as the English cancer registry. Detailed information on tumours diagnosed in England are supplied to them from multiple sources including multidisciplinary team meetings, pathology reports, hospital records and death certificates.[173,189] The NCRAS claim a near 100% case ascertainment.[189] Following collection, data undergoes a rigorous quality assurance process to maximise accuracy. Major errors, such as an invalid diagnosis date, are only found in around 0.1% of records.[173] As well as the date of diagnosis and details on how the cancer was diagnosed (e.g. histological, radiological or death certificate), detailed information on the type of cancer is collected. This includes the tumour site, recorded using the International Classification of Diseases 10th revision (ICD10) codes, and tumour morphology and behaviour, recorded using ICD-O codes.[51] The stage at diagnosis for ovarian cancer is recorded using the Tumour Nodes Metastasis (TNM) and/or the FIGO staging systems, which are essentially identical.[52,190] Although NCRAS data is updated and released regularly, there is a lag time for linkage to the CPRD and at the time the dataset for this thesis was obtained, linkage was only available up to the end of December 2015.

2.2.3 HES APC

HES APC comprises data collected during patient admissions to NHS Hospitals in England. Data on over 270 variables are available within the HES APC, including demographic information and clinical information on diagnoses and procedures.[191] CPRD linkage is available to all HES APC variables or to specific subsets of variables. For this thesis I requested data on a very limited HES APC subset of variables, in order to obtain ethnicity information. HES APC classifies ethnicity as

one of 11 categories, which map to the 2001 census.[192] The ethnicity recorded between admissions for a given patient can vary within the HES APC. NHS Digital provide a single ethnicity for each patient which represents the most common ethnicity for that patient across all episodes within HES APC.

2.2.4 Small area level data

Various geographical deprivation metrics for CPRD patients in England can be obtained as a linked dataset. Townsend score is an area level measure of material deprivation calculated on the basis of level of unemployment, car ownership, house ownership and house overcrowding in an area.[193] Areas are ranked on the basis of these variables and categorised to create a 5-level score in which high scores indicate greater deprivation. Originally calculated for the 1981 census, it has since been recalculated for the 2001 census.[194] Patients within the CPRD are linked, on the basis of postcode, to small geographical areas with an average population of around 1,600 and 5-level Townsend scores derived using 2001 census data.[194]

2.3 Regulatory approval and data delivery

Access to CRPD data and linked datasets is governed by the Independent Scientific Advisory Committee (ISAC) for the Medicines and Healthcare products Regulatory Agency (MHRA). With input from Prof Fiona Walter, Prof Willie Hamilton, Prof Gary Abel and Mrs Margaret Johnson (a patient representative), I developed a protocol covering the primary research presented in this thesis. This was submitted to ISAC in July 2018 and approved, without the need for revision, in August 2018 (protocol number 18_184, **Appendix B**). However, the preparation of my linked dataset took over five months and it was not delivered to me until the 30th of January 2019. I downloaded the zipped dataset files from CPRD using a single use password. All data were stored and analysed on a secure server in Cambridge, in accordance with CPRD terms of use and data protection legislation. I was the only researcher with access to the raw data files.

I submitted a minor amendment to ISAC on the 2/7/2019 and obtained permission to alter the primary clinical outcome definition and follow-up period (**Appendix B**). I made the decision to

make these changes prior to data analysis and I outline the rationale for these choices in **Sections 2.5** and **2.6**.

2.4 Baseline Cohort definition

The baseline cohort was defined by the presence of a code for CA125 within the CPRD between the 1st of May 2011 and the 31st of December 2014. The first CA125 test during this period was selected as the index test. Codes used to identify CA125 entries are shown in **Appendix C**. The 1st of May 2011 was chosen as a start date as this fell just after the introduction of NICE CG122 which recommended CA125 use in primary care.[114] The 31st of December 2014 was chosen to allow a follow-up period of 1 year in NCRAS for all women, and is further discussed in **Section 2.6**. Other inclusion criteria were:

- Patient flagged as Acceptable for research (CPRD quality metric)
- Registered at a practice which was UTS at the index test date (CPRD quality metric)
- Patient aged ≥18 years at index test date
- No CA125 test in the 12 months prior to index test date
- No ovarian malignancy recorded in NCRAS data on or prior to the index test date

As is recommended by the CPRD, I applied the Acceptable patient and UTS practice criteria to maximise the quality and completeness of data used in this thesis. Women with a CA125 test in the 12 months before the index test date were not included, as there is evidence that repeat blood tests in primary care may have a higher PPV than initial tests.[175] While repeat testing and trends in CA125 results are of research interest, they are not the focus of this thesis. Women with previous ovarian cancer were not included in the cohort as I was interested in the detection of primary rather than recurrent ovarian cancer.

2.5 Justification of data source for cancer diagnoses

My intention, at the time of ISAC protocol submission, was to use codes for ovarian cancer recorded in either CPRD or NCRAS data (for patients in whom linkage was available) as evidence

of an ovarian cancer diagnosis. However, a publication by Badrick *et al*. in October 2018 (after the submission of the ISAC protocol) highlighted potential pitfalls to this approach.[195]

Although many studies have used CPRD to identify cancer cases, NCRAS is generally considered the gold standard data source for cancer epidemiological studies as it collects diagnostic information from multiple sources, uses a hierarchical approach when determining date of diagnosis, records details on both the site and morphology of tumours in a standard and precise categorical format (ICD codes) and undertakes rigorous automated and manual quality control.[173] By contrast, the CPRD relies on manual coding of diagnoses by GP practice staff and it is not possible to determine the source of diagnostic information (e.g. histology or death certificate), so a hierarchical approach cannot be applied. No systematic quality controls are undertaken to ensure accuracy of CPRD diagnostic codes and diagnosis dates. Rather than ICD codes, cancer diagnoses are recorded within CPRD using Read codes, which are often so non-specific that they could refer to a benign or a malignant lesion (e.g. Read code: BB81.14, Read term: "[M]Ovarian serous tumour") or to a lesion of the ovary or another site entirely (e.g. Read code: B4….00, Read term: "Malignant neoplasm of genitourinary organ").

Multiple studies have identified differences in recording of diagnoses between NCRAS data and CPRD. Arhi *et al* found that around 10% of cancers recorded in CPRD could not be confirmed in NCRAS data while Boggon *et al* reported the figure as 17%.[196,197] Boggon *et al* also reported that 6% of cancer cases recorded in NCRAS data were missing from CPRD. A recent study, by Strongman *et al,* found that 15% of ovarian cancers recorded within the CPRD (2000-2014) could not be confirmed using a 'gold standard' algorithm which drew information from multiple data sources including NCRAS.[198] Badrick *et al* highlight that relying on CPRD data to identify outcomes could result in misclassification bias.[195] In diagnostic accuracy studies, particularly where the outcome is relatively rare, this has the potential to markedly affect estimates of test accuracy. Given these issues, instead of using both CPRD and NCRAS codes, as proposed in the original protocol, I restricted the study sample to patients who were eligible for NCRAS linkage and relied solely on NCRAS data to identify ovarian cancer diagnoses. This ensured that the same data

source, and the one which is likely to be most accurate in terms of cancer recording, was used to identify outcomes for all women. This decision was taken prior to data analysis and an amendment was approved by ISAC (**Appendix B**).

I used ICD10 codes consistent with the thesis definition of ovarian cancer as "an invasive or borderline tumour arising from the ovaries, fallopian tubes or peritoneum" to identify ovarian cancers in the dataset (**Table 2.1**).

Table 2.1. ICD10 codes meeting the thesis definition of ovarian cancer.

Code*	Description
C56	Malignant neoplasm of the ovary
C57.0	Malignant neoplasm of the fallopian tube
C48.1	Malignant neoplasm of specified parts of the peritoneum
C48.2	Malignant neoplasm of the peritoneum, unspecified
D39.1	Neoplasm of uncertain or unknown behaviour of the ovary

*All sublevel codes were included. E.g. C56.1 ("Malignant neoplasm of the right ovary").

2.6 Justification of follow-up period

A 'perfect' diagnostic accuracy study necessitates performing a gold standard test on all patients at the same time as, or shortly after, the test under evaluation.[170] This was not possible in this thesis, which instead relied on information collected on cancers diagnosed as part of standard clinical care. The length of follow-up period after CA125 testing was therefore important – it needed to be long enough to include cancers diagnosed as part of usual care, but short enough to minimise the risk of capturing incidental ovarian cancers, which were not present when the index CA125 test was performed, but which developed and were diagnosed later in the follow-up window.

While there were no comparable primary care studies on CA125 to help guide this decision, studies have evaluated the symptoms of ovarian cancer reported in primary care before diagnosis (case-control studies) and evaluated the performance of symptom-based tools in primary care. These studies have variously used follow-up periods of 6 months (a USA study),[172] 12 months,[91] and 24 months.[54] A recent study (in which follow-up was limited to 12 months) found that the median diagnostic interval (time from first presentation in primary care to diagnosis) for ovarian cancer in England was 56 days, but that 10% of patients had intervals between 281-365 days.[82] This indicates that a follow-up period of 12 months is likely to pick up most, if not all, ovarian cancers present at the time of CA125 testing. Given the low incidence of ovarian cancer in the general population (**Chapter 1**), I believe that it is unlikely that significant numbers of incidental ovarian cancers arose and were diagnosed in the study cohort within 12 months after the index CA125 test.

I considered applying a shorter follow-up period e.g. 6 months. However, this would have risked excluding a significant proportion of ovarian cancers, based on data from Price *et al.*[82] I also hypothesised that in women with ovarian cancer, those with normal CA125 results (false negatives) were likely to take longer to diagnose after testing than those with abnormal results (true positives). If true, a 6-month follow-up period would have preferentially excluded women with normal CA125 results, introducing bias into estimates of test diagnostic accuracy.

Ultimately, I included ovarian cancers diagnosed within the 12 months after the index CA125 date (as recorded in NCRAS data) as my primary clinical outcome. I submitted and gained approval for an ISAC protocol amendment, as the original protocol applied a follow-up period of 24 months (**Appendix B**). As NCRAS linkage was only available up to 31/12/2015 when the data for this thesis was acquired, only patients with an initial CA125 test in the CPRD on or before 31/12/2014 were included, to allow 1-year follow-up of all patients in NCRAS data.

2.7 Data supplied by CPRD

The CPRD provided records of 135,564 research Acceptable patients with a code for a CA125 test within their primary care record between 01/05/2011 – 30/9/2016. Linked NCRAS and HES APC data was provided for 72,182 women and Townsend score for 72,191 women.

CPRD data was supplied in 51 sub-files which mapped to the file types outlined in **Box 2.1**. The files contained millions of entries (the clinical and test files alone consisted of over 100 million coded entries), the majority of which were of no relevance to my research. Prior to analysis, significant data preparation was required.

Raw NCRAS data was supplied in 2 files. The first included the date of diagnosis, site of diagnosis (ICD10), morphology and behaviour (ICD-O) codes for each tumour (recorded against a tumour ID) for each patient (recorded against a patient ID) identified by the NCRAS between 1990-2015. The second file provided further data on stage for ovarian cancers diagnosed between 1/05/2011 – 31/12/2015.

A Small Area Levels data file containing numerical 5-level Townsend score (recorded against patient ID) was provided as was a HES APC file containing several variables including an 11-category ethnicity variable (recorded against patient ID). The preparation and use of Townsend score and HES ethnicity data are described in **Chapters 4** and **6** respectively.

A linkage file, provided by the CPRD, indicated which patients had data from each linked data source (recorded against patient ID).

I imported all files, which were supplied in text format, into Stata format (.dta) and performed all data preparation and analyses using Stata software (version 15.1).

- **Patient file**

Patient data including unique patient identifier and demographics e.g. sex, year of birth

- **Practice file**

Data on GP practices e.g. unique identifier, geographical region

- **Staff file**

Data on staff performing coding e.g. practitioner type

- **Consultation file**

Data on consultation events e.g. type of consultation, date of consultation

- **Clinical file**

Large file containing data on a wide variety of clinical variables e.g. coded symptoms, diagnoses

- **Additional clinical file**

Further clinical information on certain clinical codes e.g. smoking status

- **Referral file**

Data relating to referrals e.g. date, specialty

- **Immunisation file**

Data on immunisations e.g. date, type

- **Test file**

Data on tests including laboratory investigations e.g. type, test level/result, test units

- **Therapy file**

Data on prescriptions issued e.g. medication code, dose

Box 2.1. **CPRD file types.**

2.8 Data preparation

In this section, I describe the preparation of key variables used to select the baseline cohort in addition to several variables which, although not used in cohort selection, were utilised across multiple subsequent chapters of this thesis.

2.8.1 CPRD: CA125 test preparation

The CPRD supplied data for both NCRAS linkage eligible and ineligible patients. I used the linkage eligibility file to exclude patients without NCRAS linkage. Patients with a CA125 medcode between 01/05/2011 – 31/12/14 in their test file were then selected (69,394 CA125 codes for 55,487 women).

It was important to examine CA125 records to ensure that the data was complete and without obvious errors or quality issues. So, I inspected the CA125 levels (recorded as numerical values), the units (coded as numbers which could be deciphered using additional 'look-up' files) and the laboratory upper reference limits (recorded as numerical values) for CA125 entries within the Test file.

I identified 3,596 CA125 entries which had no CA125 level recorded. These may represent tests which were not performed or duplicate requests (in some cases multiple test entries on the same day were present). Where recorded, CA125 levels ranged widely from 0 to 32,470. High values were not unexpected as, although CA125 levels are reported to be below 35 U/ml in around 99% of healthy women,[118] patients with advanced ovarian cancer can have levels in the tens of thousands.

By convention, CA125 is measured and reported in international units (IU), often simply referred to as units (U) per millilitre (ml) or the equivalent international kilounits per litre (IKU/L) or kilounits per litre (KU/L). On examination of the units for CA125 entries I found that 5% did not have a unit recorded and 3% were associated with units not used to measure CA125, including mmol/L (commonly used to measure blood glucose) and μg/L (used for multiple blood tests

including carcinoembryonic antigen and ferritin). I explored the units by the 10 English geographical regions within CPRD and by GP practice to determine whether particular labs or practices were consistently using incorrect CA125 units, but errors were not restricted to a few regions or practices. For example, the erroneous mmol/L occurred for CA125 entries from all ten regions and more than 100 practices. The median CA125 level for entries in standard units was 12 and the mean 39, compared to a median of 16 and a mean of 163 for levels with no unit or the incorrect unit. 2% of CA125 values associated with no unit or an erroneous unit were recorded as a having a level of 0 which is, physiologically, highly unlikely and which was not found for any of the 63,678 entries which possessed the correct unit.

The CA125 upper reference limit of 'normal', derived by Bast *et al* in 1983 and subsequently used by NICE in their guidelines on ovarian cancer detection, is <35 U/ml. This is the reference value I use in this thesis. I therefore did not employ the laboratory upper reference limits provided in CPRD data for analysis, but I did examine them as a possible flag for data quality and completeness. I found that 7,833 entries had no upper reference limit recorded and a small number (115) had a clearly erroneous upper reference limit ("0", "5", "245", "420" or "455").

Test result are generally automatically transferred from labs to GP practices and the information is then recorded within the Test file in CPRD data – this should maximise quality and minimize errors. Despite this, there were clear issues with both units of measurement and upper thresholds for some CA125 records. It is possible that CA125 entries with missing or erroneous units or upper thresholds represent true CA125 tests results, with errors being introduced at some point during transfer from lab to GP practice to CPRD. However, these entries could equally represent other test results erroneously coded under a CA125 test code. It is also technically possible for CA125 test results entered manually within a practice to appear within the Test file and errors or omissions could occur on manual data entry. Given the uncertainty surrounding these entries, and in order to maximise CA125 data quality, I excluded entries recorded in incorrect units or where no unit was given and entries with an incorrect upper threshold or where no threshold was given. I performed sensitivity analyses in **Chapter 3** in which I included entries:

a) regardless of unit and b) regardless of upper reference limit. CA125 entries where the CA125 level was missing were also excluded as a test level was essential to perform the primary research described in this thesis.

I selected the first valid CA125 entry for each patient during the study period. A small number of patients had more than one CA125 entry on the initial test date, so the mean was taken to give a single CA125 value (index test result) on the initial test date (index test date). I then excluded patients with CA125 tests in the year before the index test date.

2.8.2 CPRD: Age at index test date

Only the year of birth is given in CPRD data to protect confidentiality. I assigned all patient the same day (1st) and month (July) of birth which was used to calculate patient age in years at index test date. This derived age was used to exclude patients <18 years old at the index test date and then in subsequent analyses.

2.8.3 NCRAS: ovarian cancers

I identified entries with an ICD10 topography code for ovarian cancer within NCRAS data and excluded patients with an ovarian cancer code dated on, or prior to, the index test date. The first code for ovarian cancer occurring in the 12 months following the index test date was identified and formed the primary clinical outcome.

2.8.4 Baseline cohort

The final baseline cohort consisted of 50,780 women. **Figure 2.1** illustrates the application of exclusion criteria.

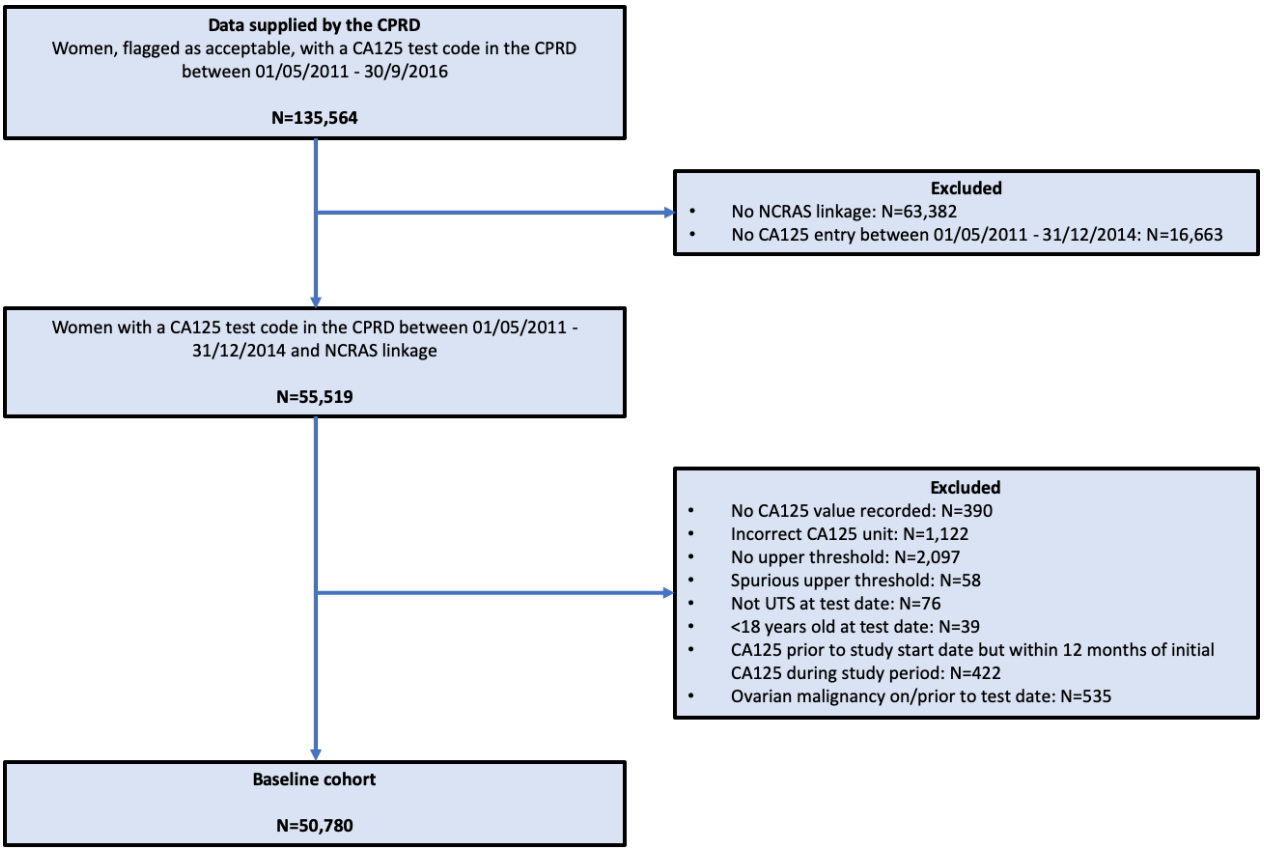


Figure 2.1. Flow diagram illustrating the selection of patients for the baseline cohort.

2.8.5 Additional NCRAS variables: morphology, behaviour and stage

While not required to select the baseline cohort, knowledge of the behaviour, morphology and stage of ovarian tumours was essential to perform the research described in subsequent chapters. The coding and preparation of these variables is best described together and is therefore presented in this section, rather than in later chapters, to minimise duplication.

Morphology and behaviour

Tumours within linked NCRAS data are coded using 2 coding systems: ICD10 for tumour 'topography' (site) and ICD-O for tumour 'morphology' and 'behaviour'. An example of the code format is given in **Figure 2.2**. As well as providing an indication of the tumour site, the ICD10 topography code indicates whether the tumour is of malignant / invasive, or of unknown / uncertain behaviour. The four-digit ICD-O morphology code provides details of the morphology

/ histology (the terms are used interchangeably in ICD-O and in this thesis) of tumours. Finally, a single digit denotes the ICD-O behaviour of the tumour.[**]

Although information on tumour behaviour is provided both in the ICD10 topography code and the ICD-O behaviour code, it became apparent that neither of these could reliably be used to determine tumour behaviour due to changes in coding practices. The previous ICD-O coding system (ICD-O-2),[199] recommended that borderline tumours were assigned a behaviour of "3" (malignant behaviour) whereas the most recent version of the coding system (ICD-O-3),[51] recommended coding them under "1" (unknown or uncertain behaviour). Likewise, borderline tumours were historically given an invasive ICD 10 code ("C") but are now often coded using the unknown / uncertain behaviour code ("D"). An examination of my data revealed that most, but not all, tumours were coded using ICD-O-2 and borderline tumours were coded under both ICD10 "C" and "D" codes.

[**] Information on tumour behaviour is provided both within the ICD10 topography code and ICD-O morphology / behaviour codes. This redundancy exists because the systems were not originally intended to be used together. The ICD-O system has a topography coding section which only provides information on site, not behaviour, but an ICD10 code is provided instead of this for CPRD-NCRAS linked data.

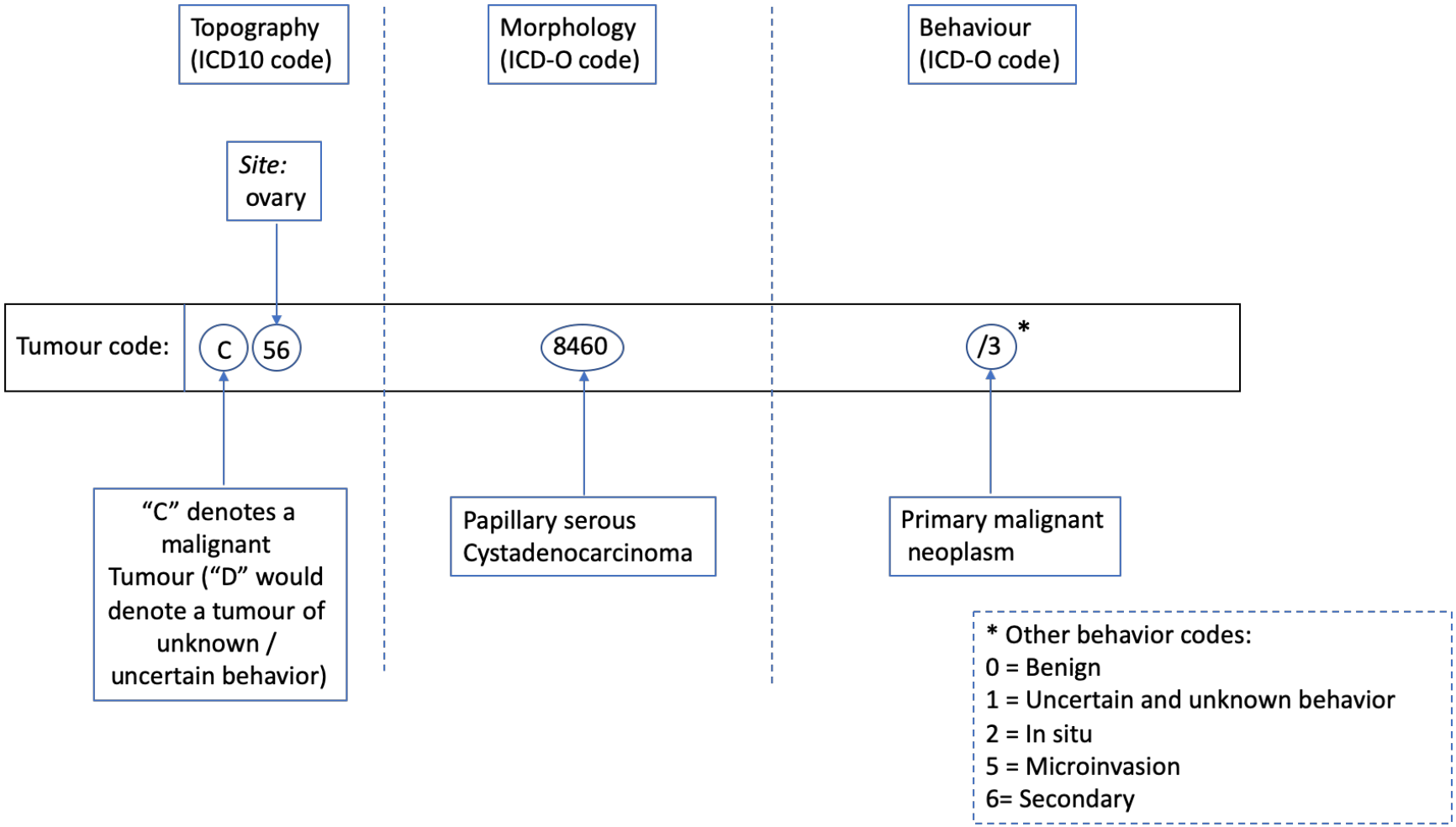


Figure 2.2. Schema illustrating the format of ICD10 and ICD-O codes using the example of "C56 8460/3", coding a "malignant papillary serous cystadenocarcinoma".

Information on tumour behaviour is also contained within the ICD-O morphology codes and, while it is more complicated and time consuming than simply relying on ICD10 codes, these can be used in conjunction with ICD-O behaviour codes to accurately determine tumour behaviour. I manually compared these codes against the ICD-O coding handbooks and categorised tumours as of borderline or invasive behaviour. At the same time, I classified tumours on the basis of ICD-O morphology code into the histological groups outlined in **Box 2.2**. While the majority of tumours could be classified by histological sub-type, some codes were non-specific e.g. "Neoplasm malignant" and "Carcinoma NOS", necessitating the histological groups: "Unknown" and "Unknown epithelial". It was not possible to further classify the serous tumour group into HGS and LGS as information on grade was recorded for a minority of tumours.

- Epithelial
  - Serous
  - Mucinous
  - Endometrioid
  - Clear cell
  - Other
  - Unknown
- Non-epithelial
- Unknown

Box 2.2. Histological classification of ovarian tumours applied in this thesis.

My provisional behaviour and histology classification lists were reviewed by Dr Brian Rous, an experienced pathologist at the NCRAS who specialises in gynaecological tumours. He provided clarification on obscure codes and suggested minor changes to the provisional list. The final classifications are included alongside relevant codes in **Appendix C**.

Stage

There are separate fields within linked NCRAS data for the individual T, N and M components of the TNM tumour stage to be recorded, in addition to the final TNM stage and a FIGO stage. Where the final TNM stage was recorded I accepted it as the final tumour stage. However, I noted that for a small number of tumours (n=6), staging information was provided in individual T, N or M fields but no final TNM stage was listed. In consultation with Dr Rous, I used the individual TNM fields to derive the final TNM stage for these tumours. In some cases, a FIGO stage was listed but no information provided in the TNM fields. Where no TNM stage could be determined and a FIGO stage was present, I accepted the FIGO stage as the final tumour stage (n=15).

Tumours were classified as:

- Stage I
- Stage II
- Stage III
- Stage IV
- Stage unknown

2.9 Chapter summary

In this chapter I have described the data used during my doctoral research and its initial preparation. In addition, I have outlined the methodological decisions taken when specifying my baseline cohort and preparing key variables. In the following four chapters, I present four research studies (three of which use the data described within this chapter) that address my specific thesis objectives.

Chapter 3. The diagnostic performance of CA125 for the detection of cancer in primary care: a population-based cohort study

3.1 Introduction

In **Chapter 1** of this thesis I highlight that, although CA125 is recommended in England and several other countries as a test for ovarian cancer in women with relevant symptoms in primary care, its diagnostic performance within this setting remains unknown. Our knowledge of CA125 performance comes from screening studies and studies conducted in the specialist setting, where the disease prevalence and the clinical characteristics of the population are distinct from primary care. To gain an accurate understanding of how CA125 performs within primary care it was important to evaluate it within an appropriate primary care population. As the incidence of ovarian cancer in primary care is low and CA125 is already widely used, and advocated by NICE, a prospective diagnostic accuracy study was not practical within this thesis. However, the fact that CA125 is widely used in primary care in England, and that test results are captured within the CPRD, provided an ideal opportunity to perform a cohort study using routinely collected data in order to evaluate how the test performs in real world primary care. In this chapter, I present a population-based cohort study which used data from the CPRD and the NCRAS to achieve the following thesis objectives:

i) *To evaluate the diagnostic accuracy of CA125 for a) ovarian cancer, b) non-ovarian cancers and c) all cancers combined, when used in English primary care.*
ii) *To explore variations in the diagnostic accuracy of CA125 by patient age.*
iii) *To explore the association between CA125 level and estimated probability of a) ovarian cancer and b) all cancers combined, in women undergoing CA125 testing in English primary care.*
iv) *To identify the CA125 level at which a 3% probability of a) ovarian cancer and b) all cancers combined is reached in women undergoing CA125 testing in English primary care.*

v) *To explore variations in estimated cancer probability by age and determine the CA125 level at which a 3% probability of a) ovarian cancer and b) all cancers combined is reached in different ages.*

A paper, based on the work presented within this chapter, was recently published in PLOS Medicine (**Appendix A**):

The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. Garth Funston, Willie Hamilton, Gary Abel, Emma J. Crosbie, Brian Rous, Fiona M. Walter. PLOS Medicine. 2020; 17:e1003295

3.2 Key concepts in the study design

In this section I briefly outline key concepts relating to test diagnostic accuracy, which were important in informing the study design. I then discuss my rationale for including age as a variable in analyses and my rationale for examining the diagnostic accuracy of CA125 for non-ovarian cancers.

3.2.1 Diagnostic accuracy considerations

In English primary care, a cut-off of ≥35 U/ml is used to interpret a patient's CA125 result. Below this threshold the test is deemed 'normal' and further investigation for ovarian cancer is not recommended by NICE.[114] Above this threshold the result is 'abnormal' and further investigation is recommended. The PPV of a test indicates the percentage of people with a test level above a given cut-off who have a condition of interest. It can be interpreted as the 'average' probability or risk of disease in patients with an 'abnormal' test result. However, patients with an 'abnormal' CA125 test can have CA125 values which range from 35 U/ml into the tens of thousands of U/ml. Patients with very high CA125 levels are more likely to have ovarian cancer than those with borderline high CA125 levels. The PPV provides no information about the probability of cancer associated with a *specific* CA125 level and is therefore of limited use when interpreting an individual woman's test result.

Knowledge of the PPV, and other conventional test accuracy metrics for CA125, are important in order to understand how the test performs overall at the threshold currently used in clinical

practice. These metrics have not previously been calculated for a primary care population, and one of the objectives of this study was to address this. However, I also sought to estimate the probability of cancer at specific CA125 levels, as I believe this is likely to be more useful clinically than the overall PPV. Patients interpret risk differently and previous research has shown that there is variation in the level of risk at which patients opt for cancer investigation.[200] Knowledge of the probability of cancer at their given CA125 level could help patients and their GPs interpret individual CA125 results and make informed decisions about further investigation.

NICE revised their cancer guidance in 2015, using a 'risk threshold' of ≥3% as the threshold for urgent cancer investigation in symptomatic women. But ovarian cancer guidance, including the chosen CA125 cut-off of 35 U/ml, remained unchanged.[57] In this study, I sought to calculate the CA125 level which equated to this 3% risk threshold. When developing their guidelines, NICE used PPV as the 'risk threshold', which was appropriate for symptoms given that they are effectively binary variables which can be 'present' or 'absent' (when severity and duration are not considered). However, such an approach is not appropriate for continuous variables with a wide range of values as, although the overall probability of disease above a given cut-off (i.e. the PPV) may be 3%, the probability at specific test values above that cut-off may vary considerably, as illustrated for a hypothetical test in **Figure 3.1.** Instead of calculating the CA125 cut-off that resulted in a 3% PPV, in this study I aimed to identify the *specific* CA125 level at which the probability of ovarian cancer reached 3%. I interpret this as the 3% 'risk threshold' in this thesis.

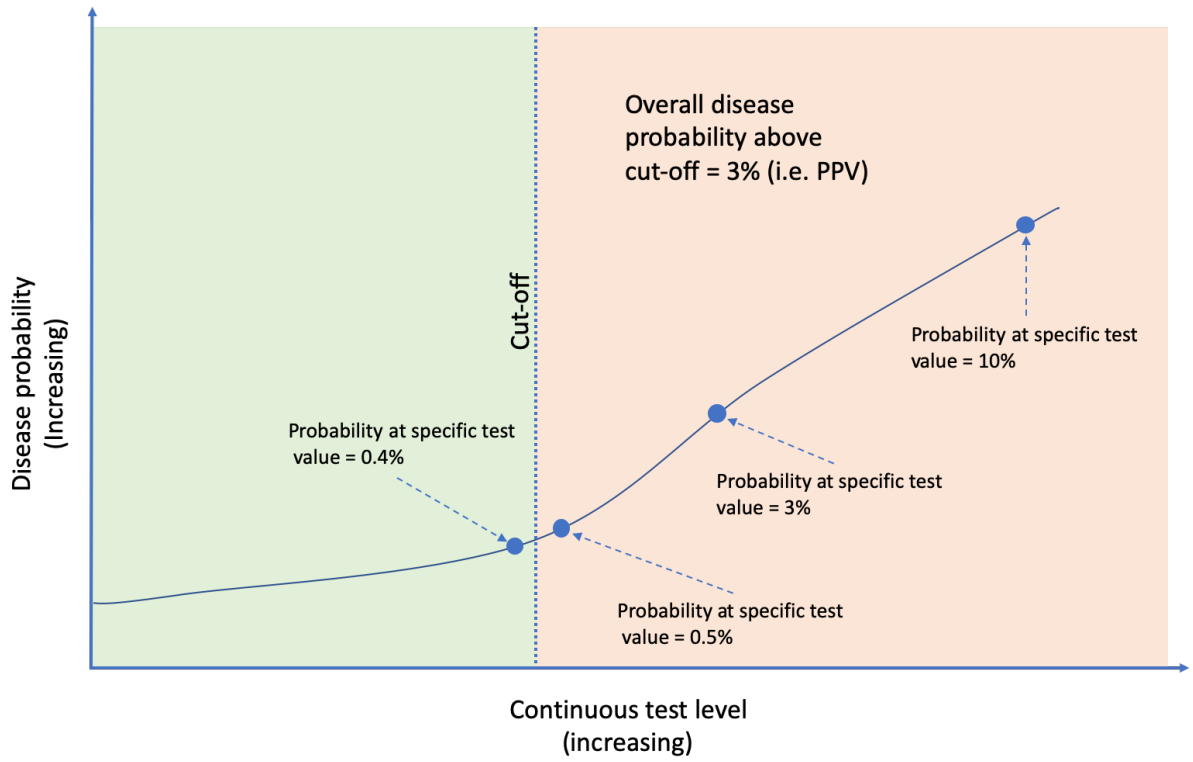


Figure 3.1. Representation of a hypothetical relationship between blood test level (continuous) and disease probability.

3.2.2 Age

As discussed in **Chapter 1,** the incidence of ovarian cancer rises markedly with increasing age; 75% of cases occur in women aged ≥50 years.[5] The low incidence of ovarian cancer in younger women is acknowledged within NICE recommendations on ovarian cancer, which state that ovarian cancer be considered in women with relevant symptoms "especially if 50 or over".[57] Screening trials, such as UKCTOCS, have gone further by only including women aged 50 or older.[154] The types of ovarian cancer which occur in younger and older women also vary. For example, germ cell ovarian cancers classically occur in younger women. In addition to cancer incidence and type, baseline CA125 levels vary by age with higher levels reported in younger groups of women.[201,202] Given this, I chose to explore the effect of age on the diagnostic accuracy of CA125 within this study. I also took account of age when estimating cancer probability for different CA125 levels.

3.2.3 Non-ovarian cancers

CA125 is only recommended as a test for ovarian cancer in primary care and the main focus of this chapter, and indeed this thesis, is on its utility in ovarian cancer detection. However,

as discussed in **Chapter 1**, CA125 levels can be elevated in the presence of other types of cancer. Most ovarian cancer symptoms are nonspecific e.g. abdominal pain, weight loss and fatigue - these are common across multiple types of cancer, including malignancies sometimes associated with elevations in CA125. This overlap in cancer presentation is demonstrated in **Table 3.1**, which shows symptoms of ovarian cancer included in NG12 which are also listed as symptoms of other CA125-elevating cancers. I postulated that a proportion of women being tested for CA125, due to a suspicion of ovarian cancer, would actually have a non-ovarian cancer and that some would have elevated CA125 levels as a result of that non-ovarian cancer. In light of this, I examined the diagnostic accuracy of CA125 for non-ovarian cancers and all cancers combined (ovarian and non-ovarian).

Table 3.1. Symptom overlap between ovarian cancer and other cancers known to cause CA125 elevations.

		Cancer					
		Ovarian	Pancreatic	Colorectal	Oesophageal	Stomach	Lung
Symptom	Abdominal pain	•	•	•	•	•	
	Loss of appetite	•					•
	Weight loss	•	•	•	•	•	•
	Fatigue	•					•
	CIBH	•	•	•			

Some ovarian cancer symptoms (urinary urgency/frequency, abdominal distension/bloating and pelvic pain) are not shown, as they are not listed as symptoms of other CA125-elevating cancers in NICE guidelines. CIBH = change in bowel habit.

3.3 Methods

3.3.1 Study design and participants

This study followed a cohort design utilising linked data from the CPRD and the NCRAS. The baseline cohort, the preparation of which is described in detail in **Chapter 2**, directly formed the study cohort. All women in this cohort were included in this study in order to maximise sample size and the precision of diagnostic accuracy estimates. To summarise, this cohort consisted of 50,780 women registered at GP practices in England who had a valid CA125 test recorded in their CPRD record between 1st of May 2011 and the 31st of December 2014. These

women had no history of ovarian cancer recorded by NCRAS on or prior to their first CA125 test within the study period (the index CA125 test) and no CA125 tests recorded in CPRD in the year prior to the index test. All patients had NCRAS linkage, allowing for 12 months follow-up in NCRAS after the index test date.

3.3.2 Rationale for not restricting the cohort by symptom codes

CA125 is only recommended as a primary care test in England for women presenting with specific symptoms. As such, I considered restricting the cohort to women with a code for any of the symptoms of possible ovarian cancer listed in NICE guidelines.[57] I chose not to do this for two reasons. First, as discussed in **Chapter 2**, symptoms are not always coded within primary care records but instead may be recorded within the free text (access to which is no longer available for research). In one large CPRD based case-control study by Price *et al*, symptoms of bladder and pancreatic cancer were recorded in free text but not coded for 38% of patients.[187] Symptoms are less likely to be coded for patients without cancer than with cancer. For example, Price *et al* found that 18% of haematuria in patients subsequently diagnosed with bladder cancer was only recorded in free text (not coded) compared to 42% of matched controls. Although the reasons for this are unclear, it may be that the symptoms in the cancer patients are more concerning to GPs or are more severe. Only including patients with a relevant symptom code before their CA125 test could therefore have biased the study results by preferentially selecting cancer patients and patients with more severe symptoms (who may have more severe disease and thus higher CA125 levels).

Secondly, even if it had been possible to definitively identify patients with symptoms of possible ovarian cancer listed in NICE guidelines, GPs frequently don't adhere to guidelines.[161] While there has never been an ovarian cancer screening program in England and it is unlikely that significant numbers of CA125 tests are performed in asymptomatic individuals in general practice, it is possible that this does occasionally occur, or that some CA125 tests are requested due to symptoms not listed in NICE guidelines. Instead of focusing on the performance of CA125 when used in concordance with guideline recommendations, I instead sought to determine its performance when used in real world primary care.

3.3.3 Handling of CA125 data

The preparation of index CA125 test data is described in **Section 2.8.1**. After visually inspecting this data, I generated two variables for subsequent analyses. First, I categorised women into two groups: CA125 'normal' (<35 U/ml) and 'abnormal' (≥35 U/ml). This variable was used to assess diagnostic accuracy at the conventional threshold (**Section 3.3.9**). Second, as CA125 level data (as a continuous variable) was highly right-skewed, I log transformed it. CA125 level was also mean centred prior to analysis. This variable was used in all regression analyses described in **Section 3.3.9**.

3.3.4 Handling of patient age

The preparation of age at index test date is described in **Section 2.8.2**. After visually inspecting data on age, I used it to categorise women into '<50 years' and '≥50 years' age groups. These were chosen as the most appropriate age groups as: more than three quarters of ovarian cancers are diagnosed in women ≥50 years of age; age 50 is the standard age used to select women for screening and early detection studies;[55,154] it is close to the mean age of menopause in the UK (51 years);[203] and is highlighted in NICE guidelines on ovarian cancer detection.[57] All analyses described within this chapter were first performed in the overall cohort then repeated by age group.

On submission of the research paper on this study, a peer reviewer suggested that, as well as calculating the estimated cancer probability at specific CA125 levels in the two age groups, I perform analyses including age as a continuous variable in order to calculate the estimated probability of cancer at specific CA125 levels for specific ages. Although such an approach was not prespecified, it was in keeping with the study objectives. A variable consisting of mean centred age in years (as a continuous variable) was therefore generated for use in this analysis.

3.3.5 Primary outcome

The primary clinical outcome in this study was the diagnosis of ovarian cancer, as recorded using ICD10 codes in NCRAS data, in the 12 months following the index CA125 test (further described in **Sections 2.5** and **2.6**).

3.3.6 Secondary outcome

The secondary clinical outcome in this study was the diagnosis of any form of cancer except non-melanoma skin cancer, as recorded using ICD10 codes in NCRAS data, within 12 months of the index CA125 test. Where a patient has both an ovarian cancer and another type of cancer, I considered that any CA125 elevations are much more likely to be due to the ovarian cancer. Therefore, I only considered non-ovarian cancers in women without ovarian cancer. I identified the first non-ovarian ICD10 code recorded in the 12 months following the index test date. In cases where women had more than one non-ovarian cancer diagnosis on the same day, I used the following approach to select a cancer based on behaviour:

1) Malignant
2) Unknown or uncertain behaviour
3) In situ

In the few instances where the cancers were of the same behaviour type (n=4), I selected a diagnosis arbitrarily. I used a pragmatic approach to categorise tumours by anatomical site on the basis of ICD10 codes. Most types of in situ cancers do not cause symptoms and therefore are unlikely to have been the trigger for CA125 testing. Therefore, in situ lesions were grouped separately from tumours of malignant behaviour and those of unknown behaviour. Tumours could generally be classified according to top order ICD10 codes but in a few instances lower order codes had to be examined. For example, D37 represents “neoplasm of uncertain behaviour of the oral cavity and digestive tract” and was categorised using lower order codes e.g. D37.1 (“neoplasm of uncertain behaviour of the stomach”) as an upper gastrointestinal (GI) cancer and D37.4 (“neoplasm of uncertain behaviour of the colon”) as a lower GI cancer. Occasionally, codes were too non-specific to be grouped on the basis of topographical site e.g. D48 (“neoplasm of uncertain behaviour: other”) and these are reported individually.

Final classifications were reviewed by Dr Brian Rous (consultant pathologist at the NCRAS) who agreed with my assignments.

In this thesis when referring to cancers other than ovarian I use the term 'non-ovarian cancer'. Where I discuss the combined non-ovarian and ovarian cancer groups I use the term 'all cancers'.

3.3.7 Sub-analysis: Invasive ovarian cancer

The definition of ovarian cancer in this thesis included both borderline and invasive tumours. As discussed in **Chapter 1**, Borderline tumours are treated collectively with invasive tumours in NICE recommendations on CA125 testing,[57] are staged using the same systems (FIGO or TNM), require surgical management and can relapse.[204] Although their timely detection in symptomatic women is important, their prognosis is good even if detected late, in contrast to invasive ovarian cancers.[204] To explore how the diagnostic accuracy of CA125 would differ if borderline tumours were not considered, I performed a sub-analysis in which invasive ovarian cancer formed the outcome.

3.3.8 Descriptive outcome: ovarian morphology

The morphology of ovarian cancers diagnosed within the cohort was identified from NCRAS data (**Section 2.8.5**) and described by age group (<50 years and ≥50 years).

3.3.9 Statistical analyses

I applied two distinct methodological approaches to analyse data in this study. First, I calculated conventional measures of CA125 diagnostic accuracy for a prespecified threshold. Second, I treated CA125 as a continuous variable, explored the estimated probability of cancer at specific CA125 levels and identified CA125 levels equating to a 3% probability of cancer. I describe each approach separately.

Diagnostic accuracy for a prespecified cut-off

I used the *DIAGT* program within Stata to calculate the following CA125 diagnostic test accuracy metrics, and associated 95% CI,[205] applying the ≥35 U/ml cut-off:

- *Sensitivity*: True positives / (true positive + false negatives)
- *Specificity*: True negatives / (true negatives + false positives)
- *Positive Predictive Value (PPV)*: True positives / (True positives + false positives)
- *Negative Predictive Value (NPV)*: True negatives / (False negatives + true negatives)

These metrics were calculated for:

- Ovarian cancer
- Invasive ovarian cancer
- All cancers

And, after excluding patients with ovarian cancer:

- Non-ovarian cancer

All analyses were repeated by age group (<50 and ≥50 years).

In **Chapter 2,** I described how CA125 entries associated with a) no, or incorrect, units and b) no, or incorrect upper, reference ranges, were dropped during baseline cohort preparation. To examine whether these decisions had an impact on my results, I performed sensitivity analyses, re-calculating the diagnostic accuracy of CA125 for ovarian cancer when these CA125 entries were not dropped.

Area under the receiver operator characteristic curve (AUC)

I used the *roctab* command in Stata,[206] to calculate the AUC of CA125 for each outcome and repeated this by age group. AUC is a measure of test *discrimination* i.e. the ability of a test to distinguish those with a condition from those without a condition.[207] AUC is calculated after constructing a receiver operator characteristic (ROC) curve, in which the sensitivity of a test is plotted (on the y-axis) against 1-specificty of a test (on the x-axis) at all test values (i.e. all potential cut-offs). The AUC therefore gives an indication of overall test performance rather than test performance at a particular threshold of interest. A test with no discrimination ability would have an AUC of 0.5 while a test with perfect discrimination ability would have an AUC of 1.

Cancer probability by CA125 level

I used logistic regression analysis to examine the relationship between log CA125, as a continuous variable, and ovarian cancer diagnosis.

I explored the nature of the relationship between log CA125 level and ovarian cancer, by introducing polynomial terms into the regression model, and found it to be non-linear. To account for this, I used restricted cubic splines.[††] In restricted cubic splines the range of the continuous variable is essentially divided into a series of segments or 'splines' (piecewise cubic polynomials). Adjacent splines must be continuous at specific points known as 'knots' but between these knots there is great flexibility in the shape which they can take. The splines are restricted to be linear in the upper and lower tails, which adds stability to the model. **Figure 3.2** illustrates some of the key features of restricted cubic splines.

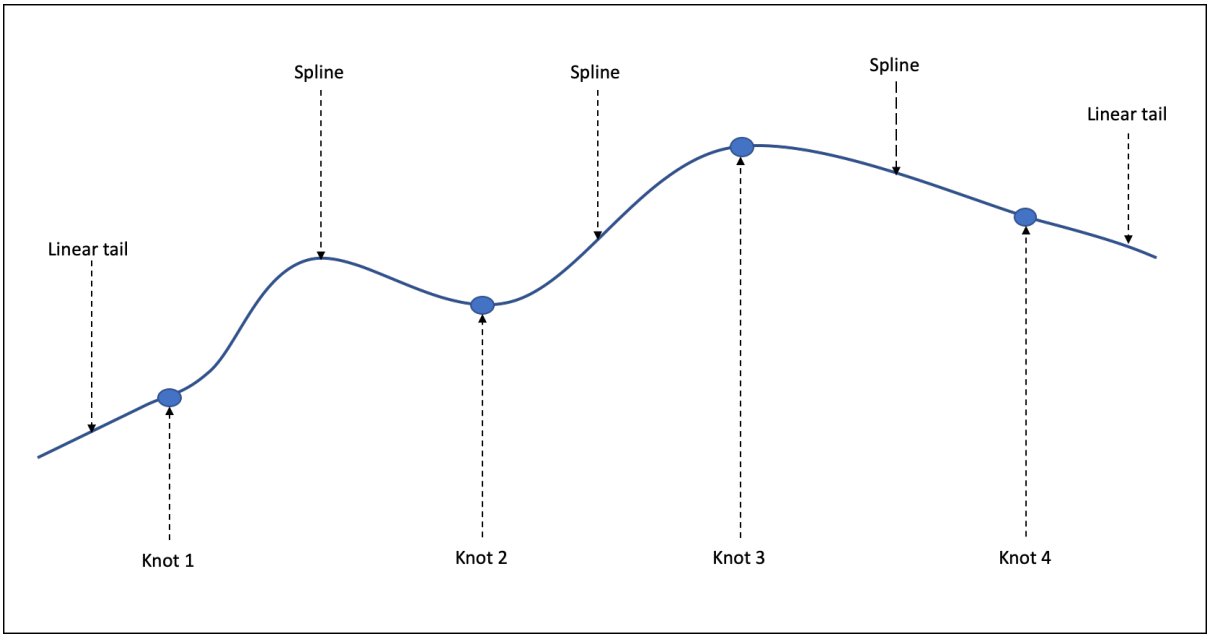


Figure 3.2. Key features of restricted cubic splines.
This image was not constructed using real data and is presented simply for purpose of illustration.

When using restricted cubic splines, two choices must be made: what number of knots to use and where they should be placed. As recommended by Harrell,[208] I compared the Akaike

[††] Splines were used in the final models, rather than polynomial terms, following feedback from a statistical peer reviewer who recommended this approach.[338]

Information Criterion (AIC) for models containing three, four and five knots. The five-knot model produced the smallest AIC and so I chose to use it. I also followed Harrell's advice on knot positioning, placing them at standard, equally spaced percentiles of the marginal distribution of the variable (**Appendix D**).[208] I utilised the *mkspline* function within Stata to generate the restricted cubic splines for the regression model. I used this regression model to predict the odds of cancer for a range of CA125 levels (1-1000 U/ml), then converted these odds into probabilities, using Stata code provided by Prof Gary Abel.

All these steps were repeated using a) invasive ovarian cancer and b) all cancers, as the outcome.

Cancer probability by CA125 level and age

I repeated the logistic regression analysis for the <50 years and ≥50 years groups. I then constructed a multivariable regression model including splines for age in years, and log CA125 level, applying the same approach as described above. Five knots were included for each variable (**Appendix D**). This regression model was used to predict the odds of ovarian cancer for CA125 levels (1-1000 U/ml) in women of different ages. These odds were converted to probabilities for women aged 30, 40, 50, 60, 70 and 80 years of age.

I considered including an interaction term in my analysis, to account for any interaction between log CA125 and age. However, this term was not significant in an initial logistic regression model and so was not included in the final model.

These steps were repeated using a) invasive ovarian cancer and b) all cancers, as the outcome.

3.4 Results

3.4.1 Study cohort

The final study cohort consisted of 50,780 women. The selection of this cohort is described in detail in **Chapter 2** (summarised in **Figure 2.1**).

3.4.2 Age and CA125 distributions

The mean patient age within the cohort at index test date was 56 years (range: 18-102 years). The age distribution of the cohort is illustrated in **Figure 3.3**.

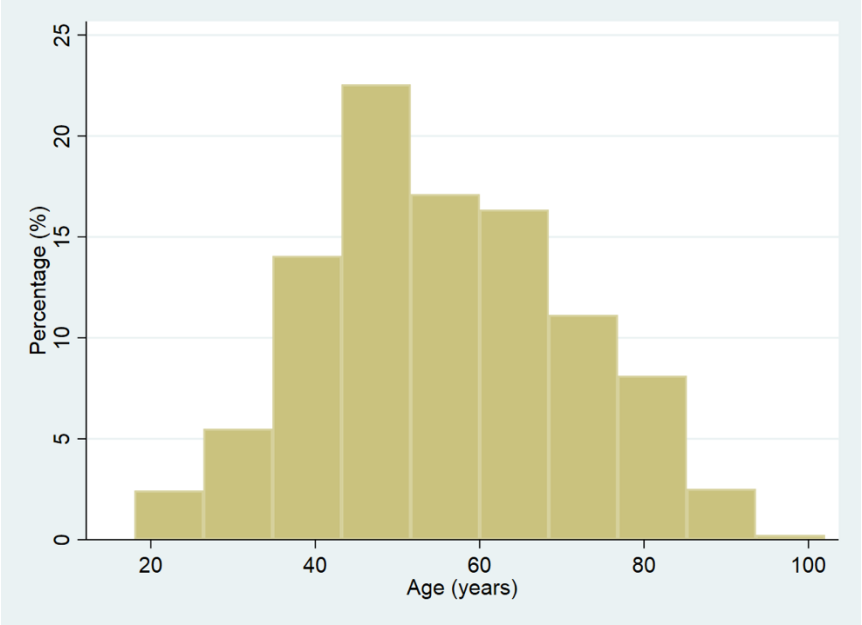


Figure 3.3. Histogram illustrating the age profile of the study cohort.

19,694 (38.8%) women were aged <50 years and 31,086 (61.2%) women were aged ≥50 years (**Table 3.2**).

CA125 levels ranged from 1 U/ml - 32,470 U/ml, with a median value of 12 U/ml, (interquartile range [IQR]: 8-18). The median value in women aged <50 years was 13 U/ml (IQR: 9-20 U/ml) and in women ≥50 years 11 U/ml (IQR: 8-17 U/ml). 3,468 (6.8%) had an index CA125 level ≥35 U/ml (**Table 3.2**). The proportion of women with CA125 levels ≥35 U/ml was slightly higher in the <50 years group (7.5%) than the ≥50 years group (6.8%).

Table 3.2. Patient numbers, incidence of raised CA125 tests (≥35 U/ml) and cancer incidence by age group.

	<50 years	≥50 years	Overall cohort
Number of patients, N	19,694	31,086	50,780
Raised (≥35 U/ml) CA125, N (%)	1,482 (7.5)	1,986 (6.4)	3,468 (6.8)
Ovarian cancers, N (%)	80 (0.4)	376 (1.2)	456 (0.9)
Non-ovarian cancer, N (%)	161 (0.8)	1,160 (3.7)	1,321 (2.6)

3.4.3 Cancer incidence

456 (0.9%) women were diagnosed with ovarian cancer in the year following the index test date (**Table 3.2**). The incidence of ovarian cancer was three times higher in the ≥50 years group (1.2%) than ≤50 years group (0.4%). In those not diagnosed with ovarian cancer (n=50,324), 1,321 (2.6%) women were diagnosed with another type of cancer in the 12 months following the index test.

3.4.4 Ovarian cancer morphology

Of the ovarian cancers diagnosed, 98 (21.5%) were borderline tumours. The proportion of malignancies which were borderline varied by age, with 50% of tumours in the <50 years group and 15.4% in the ≥50 years group being of borderline malignancy (**Table 3.3**). Serous epithelial tumours were the most common tumour type, accounting for 48.6% of invasive tumours; 42.5% in the <50 years group and 71.7% the ≥50 year. 15% of invasive tumours in the <50 years group were of mucinous epithelial origin and 12.5% were of non-epithelial origin compared to 5.3% and 2.5% respectively in the ≥50 years group.

Table 3.3. Behaviour and histology of ovarian tumours by age group (<50 years and ≥50 years).

Behaviour and Histology	<50 years	≥50 years	All ages
Invasive			
Epithelial:			
Serous	17	157	174
Mucinous	6	16	22
Clear Cell	0	17	17
Endometrioid	4	16	20
Other	2	15	17
Unknown	4	69	73
Non-epithelial	5	8	13
Unknown	2	20	22
Borderline			
Borderline	40	58	98
Total	***80***	***376***	***456***

3.4.5 Diagnostic accuracy

In this section I present the diagnostic accuracy metrics for ovarian cancer, invasive ovarian cancer and all cancers for the conventional CA125 cut-off (≥35 U/ml). The AUC of CA125 for each outcome in different age groups is also presented.

Ovarian cancer

The diagnostic performance characteristics of CA125, calculated after applying the standard cut-off (≥35 U/ml), are summarised in **Table 3.4**. At or above the 35 U/ml cut-off, CA125 demonstrated a PPV of 10.1% (95% CI: 9.1-11.2), an NPV of 99.8% (95% CI: 99.7-99.8), a sensitivity of 77.0% (95% CI: 72.8-80.8%) and a specificity of 93.8% (95% CI: 93.6-94.0) for ovarian cancer. The PPV, sensitivity and specificity were all higher for the ≥50 group than the <50 years group. When the outcome was restricted to invasive ovarian cancers, CA125 demonstrated a slightly lower PPV (8.8%, 95% CI: 7.8-9.8), and a higher sensitivity (84.9%, 95% CI: 80.8-88.5). The PPV, sensitivity and specificity of CA125 were all higher for invasive tumours in the ≥50 group than the <50 years group.

Overall, CA125 had an AUC of 0.92 for ovarian cancer. This was higher in the ≥50 years group (0.93) than the <50 years group (0.86). A similar pattern was seen when the outcome was restricted to invasive ovarian cancer (**Table 3.4**).

Sensitivity analyses

In a sensitivity analysis in which CA125 entries were not dropped due to incorrect or unrecorded units during baseline cohort preparation (**Section 2.8.1**), the PPV (10.2%, 95% CI: 9.2-11.3), NPV (99.8%, 95% CI: 99.7-99.8), specificity (93.8%, 95% CI: 93.6-94.0) and sensitivity (77.0%, 95% CI: 72.9-80.7) were similar to those for the main cohort. In a sensitivity analysis in which CA125 entries were not dropped due to erroneous or unrecorded upper thresholds, the PPV (10.1%, 95% CI: 9.1-11.1), NPV (99.8%, 95% CI: 99.7-99.8), specificity (93.7%, 95% CI: 93.5-93.9) and sensitivity (77.5%, 95% CI: 73.5-81.2) were again similar to values for the baseline cohort.

Table 3.4. Measures of CA125 diagnostic accuracy for ovarian cancer, invasive ovarian cancer, all cancers and non-ovarian cancer.

Cancer	Group	PPV, % (95% CI)	NPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC (95% CI)
Ovarian	All ages	10.1 (9.1-11.2)	99.8 (99.7-99.8)	77.0 (72.8-80.8)	93.8 (93.6-94.0)	0.92 (0.90-0.93)
	<50 years	3.4 (2.5-4.4)	99.8 (99.8-99.9)	62.5 (51.0-73.1)	92.7 (92.3-93.1)	0.86 (0.82-0.91)
	≥50 years	15.2 (13.6-16.8)	99.7 (99.7-99.8)	80.1 (75.7-84.0)	94.5 (94.3-94.8)	0.93 (0.92-0.95)
Ovarian: invasive	All ages	8.8 (7.8-9.8)	99.9 (99.9-99.9)	84.9 (80.8-88.5)	93.7 (93.5-93.9)	0.94 (0.92-0.96)
	<50 years	2.0 (1.3-2.8)	99.9 (99.9-100)	72.5 (56.1-85.4)	92.6 (92.2-93.0)	0.88 (0.82-0.95)
	≥50 years	13.8 (12.4-15.4)	99.9 (99.8-99.9)	86.5 (82.2-90.0)	94.4 (94.2-94.7)	0.95 (0.93-0.97)
All cancers	All ages	21.2 (19.8-22.6)	97.8 (97.7-97.9)	41.4 (39.1-43.7)	94.4 (94.2-94.6)	0.68 (0.66-0.69)
	<50 years	6.1 (4.9-7.4)	99.2 (99.0-99.3)	37.3 (31.2-43.8)	92.8 (92.5-93.2)	0.62 (0.58-0.67)
	≥50 years	32.5 (30.4-34.6)	96.9 (96.7-97.1)	42.0 (39.5-44.5)	95.5 (95.2-95.7)	0.70 (0.69-0.72)
Non-ovarian	All ages	12.3 (11.2-13.5)	98.0 (97.9-98.1)	29.1 (26.6-31.6)	94.4 (94.2-94.6)	0.74 (0.73-0.75)
	<50 years	2.8 (2.0-3.8)	99.3 (99.2-99.4)	24.8 (18.4-32.3)	92.8 (92.5-93.2)	0.70 (0.67-0.74)
	≥50 years	20.4 (18.5-22.4)	97.2 (97.0-97.4)	29.7 (27.0-32.4)	95.5 (95.2-95.7)	0.76 (0.75-0.78)

PPV, NPV, sensitivity and specificity are calculated for a cut-off of ≥35 U/ml. Accuracy characteristics for 'non-ovarian' cancer were calculated following exclusion of patients with ovarian cancer.

All cancers

Applying the ≥35 U/ml cut-off, the PPV of CA125 for all cancers combined was 21.2% (95% CI:19.8-22.6). This varied markedly by age, with a PPV of 6.1% (95% CI: 4.9-7.4) in the <50 years group and 32.5% (95% CI: 30.4-34.6) in the ≥50 years group (**Table 3.4**). The overall

specificity of CA125 for all cancers was slightly higher than for ovarian cancer but the sensitivity of the test was much lower (41.4%, 95% CI: 39.1-43.7).

Overall, CA125 had an AUC of 0.74 for all cancers. This was higher in the ≥50 years group (0.76) than the <50 years group (0.70).

Non-ovarian cancer

The incidence of non-ovarian cancers in those women without ovarian cancer (n=50,324) who had a CA125 <35 U/ml (n=42,207) was 2.0%, while the incidence in women with a CA125 ≥35 U/ml, which equates to the PPV for non-ovarian cancers, was 12.3% (95% CI: 11.2-13.5). This varied dramatically between the <50 years group (PPV: 2.8%, 95% CI: 2.0-3.8) and ≥50 years group (PPV: 20.4%, 95% CI: 18.5-22.4) (**Table 3.4**). While the overall specificity of the test for non-ovarian cancers was high (94.4%, 95% CI: 94.2-94.5%), the sensitivity was very low (29.1%, 95% CI: 26.6-31.6).

Figure 3.4 compares the incidence of non-ovarian cancer types between the <35 U/ml and ≥35 U/ml groups, after exclusion of women with ovarian cancer. All cancer types occurred more frequently in the ≥35 U/ml group than the <35 U/ml group. Pancreatic cancer was 15 times, lung cancer 13 times, uterine cancer nine times, upper GI cancer six times and lower GI cancer four times more common in the ≥35 U/ml group than the <35 U/ml group. A similar (small) proportion of women in each group were diagnosed with breast cancer (CA125 <35 U/ml: 0.3% vs CA125 ≥35 U/ml: 0.38%).

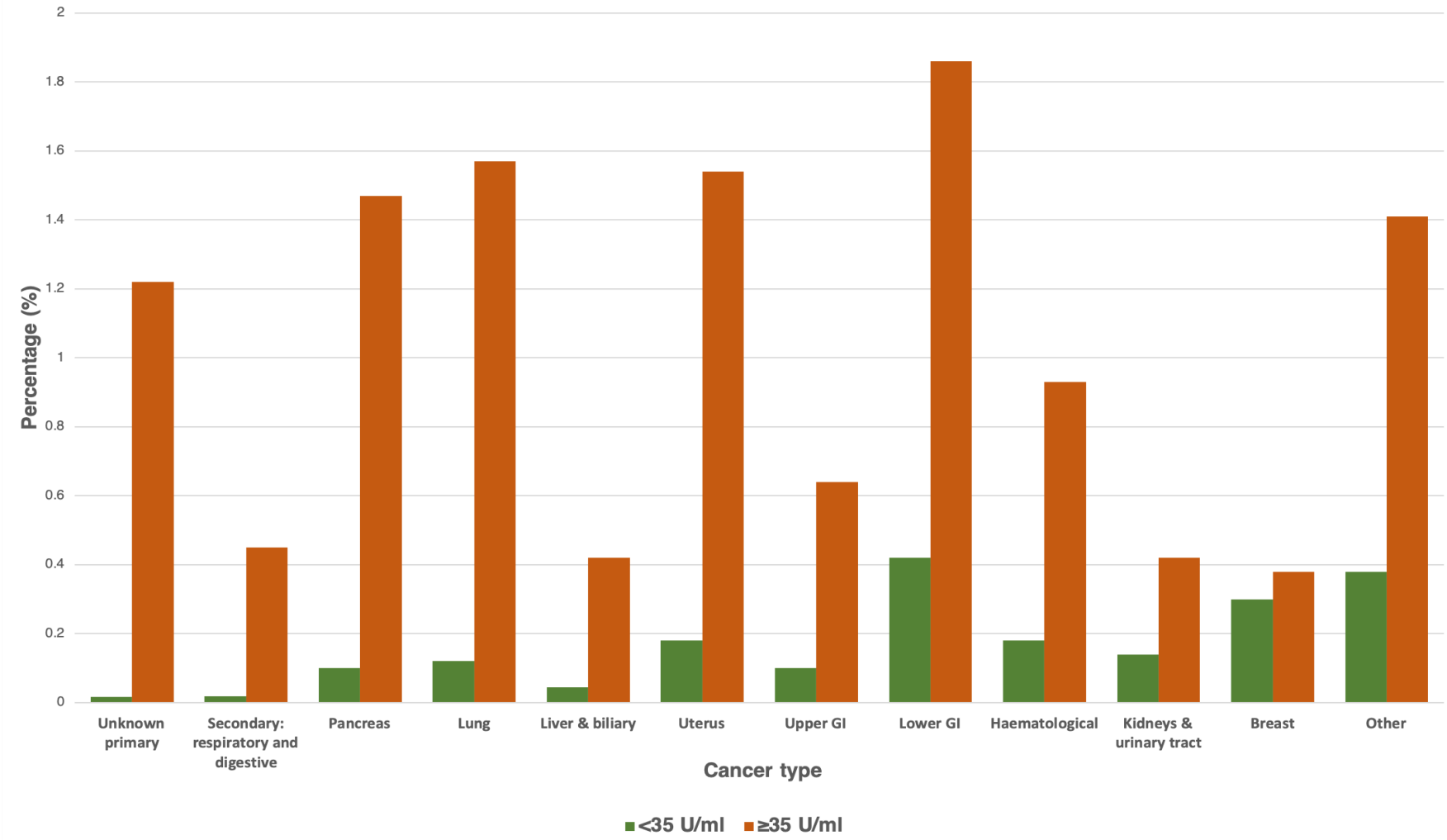


Figure 3.4. Percentages of women with non-ovarian cancer types in each CA125 group: <35 U/ml (n=47,207) and ≥35 U/ml (n=3,117). Women with ovarian cancer (n=456) were excluded. GI = gastrointestinal.

The non-ovarian tumour categories most frequently associated with CA125 levels ≥35 U/ml are shown in **Table 3.5**. Although the proportion of women with non-ovarian cancers who had CA125 levels ≥35 U/ml was 29.1% overall, this varied by cancer type. 43 women in the cohort had a cancer of unknown primary, and a greater proportion of these women had a CA125 level ≥35 U/ml (83%), than for any other cancer type (including ovarian cancer). 93 women were diagnosed with pancreatic cancer and 104 with lung cancer, almost half of whom had CA125 levels ≥35 U/ml. For the two most common categories of non-ovarian cancer - lower GI (n=255) and breast (n=154) - the proportion of women with CA125 levels ≥35 U/ml was much lower, at 23% and 8% respectively.

Table 3.5. Cancers diagnosed in women without ovarian cancer.

Cancer type (ICD10 codes)	N	N <35 U/ml (%)	N ≥35 U/ml (%)
Unknown primary (C80)	46	8 (17)	38 (83)
Secondary: respiratory and digestive (C78)	23	9 (39)	14 (61)
Pancreas (C25)	93	47 (51)	46 (49)
Lung (C34)	104	55 (53)	49 (47)
Liver, biliary (C22,C23,C24)	34	21 (62)	13 (38)
Uterus (C54,C55,D39.0)	132	84 (64)	48 (36)
Upper GI (C15,C16,C17,D37.1,D37.2)	66	46 (70)	20 (30)
Lower GI (C18,C19,C20,C21,D37.3,D37.4,D37.5)	255	197 (77)	58 (23)
Haematological (C81,C82,C83,C84,C85,C90,C91,C92,C96,D45,D46,D47)	112	83 (74)	29 (26)
Kidneys, urinary tract (C64,C65,C66,C67,D41)	78	65 (83)	13 (17)
Breast (C50)	154	142 (92)	12 (8)
Other*	224	180 (80)	44 (20)
Total	**1,321**	**937 (71)**	**384 (29)**

*Cancers with fewer than ten cases with CA125 values ≥35 U/ml.

3.4.6 Cancer probability analyses

In this section I present the cancer probabilities derived following logistic regression analyses. As the 3% cancer probability is of primary interest in this study, figures in the main text illustrate probabilities close to this threshold. Figures showing the probability for each outcome at an extended range of CA125 levels (1-500 U/ml) are included in **Appendix E**. All estimated probabilities for ovarian cancer, invasive ovarian cancer and all cancers calculated

during this study (for CA125 levels 1-1000 U/ml) have been made freely available via the University of Cambridge Repository.[209]

Cancer probability by CA125 level

Figure 3.5 illustrates how the estimated probability of ovarian cancer, invasive ovarian cancer and all cancers increased with increasing CA125 level. A CA125 level of 53 U/ml equated to a probability of 3% (95% CI: 2.6-3.5) for ovarian cancer while a CA125 level of 18 U/ml equated to a probability of 3% (95% CI:2.8-3.2) for all cancers. In a sub-analysis in which invasive ovarian cancer formed the outcome, a CA125 level of 68 U/ml equated to a 3% (95% CI: 2.6-3.5) cancer probability.

Much greater estimated cancer probabilities were noted at higher CA125 levels. For example, at a CA125 level of 500 U/ml the estimated probability of ovarian cancer was 31.6% (95% CI: 28.1-35.2) and all cancers, 60.6% (95% CI: 56.9-64.3) (**Appendix E**).

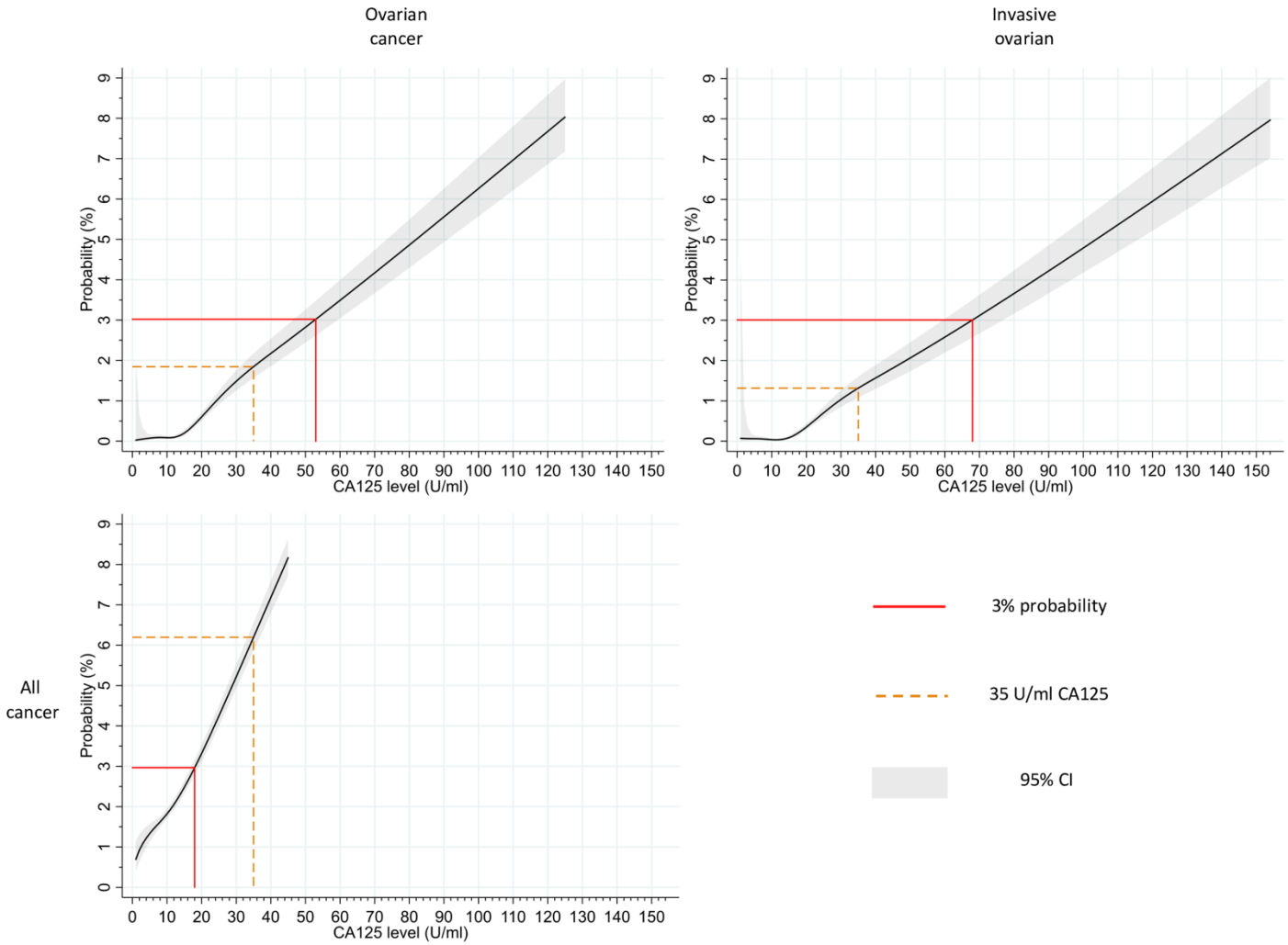


Figure 3.5. Relationship between CA125 level and estimated probability of ovarian cancer, invasive ovarian cancer and all cancers.

Cancer probability by CA125 level and age group

Figure 3.6 shows the probability of ovarian cancer derived by repeating the logistic regression analysis by age group (<50 years and ≥50 years). This analysis revealed that the 3% ovarian cancer probability threshold was reached at a much lower CA125 level in the ≥50 years group (39 U/ml) than <50 years group (89 U/ml). At equivalent CA125 levels, the estimated probability of ovarian cancer was always greater in women aged ≥50 years than those <50 years.

A similar pattern was seen when analysis was repeated by age group for invasive ovarian cancer and all cancers (**Appendix E**). For all cancers, a CA125 of 47 U/ml in women under 50 years and 14 U/ml in women ≥50 years equated to a 3% probability.

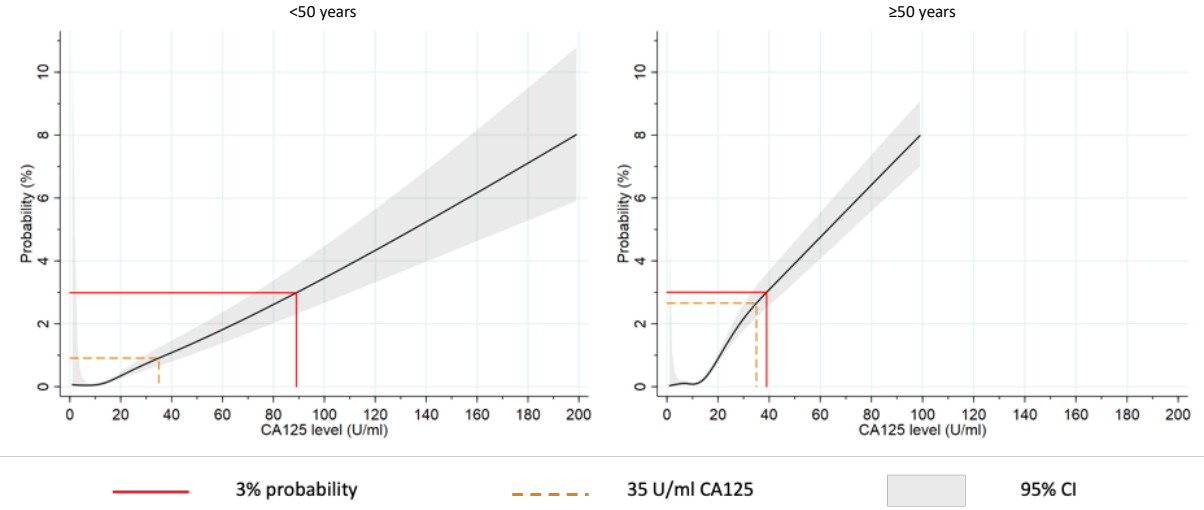


Figure 3.6. Relationship between CA125 level and estimated probability of ovarian cancer in women <50 years and ≥50 years of age.

Cancer probability by CA125 level and age in years

Figure 3.7 illustrates the relationship between CA125 level and the estimated probability of ovarian cancer at the specific ages (30, 40, 50, 60, 70 and 80 years), derived from a logistic regression in which splines for both CA125 and age in years were included. The CA125 level equating to an estimated 3% ovarian cancer probability was 60 U/ml at age 30, increasing to 104 U/ml at age 40 before falling through ages 50 and 60 to a low of 32 U/ml at age 70. This figure rose to 43 U/ml at age 80.

Similar age trends were noted when the analysis was repeated for invasive ovarian cancer and all cancers (**Appendix E**). For all cancers, a CA125 of 50 U/ml equated to a 3% estimated cancer probability at age 40, while the 3% threshold was reached at a CA125 level of only 7 U/ml at age 70 and 80.

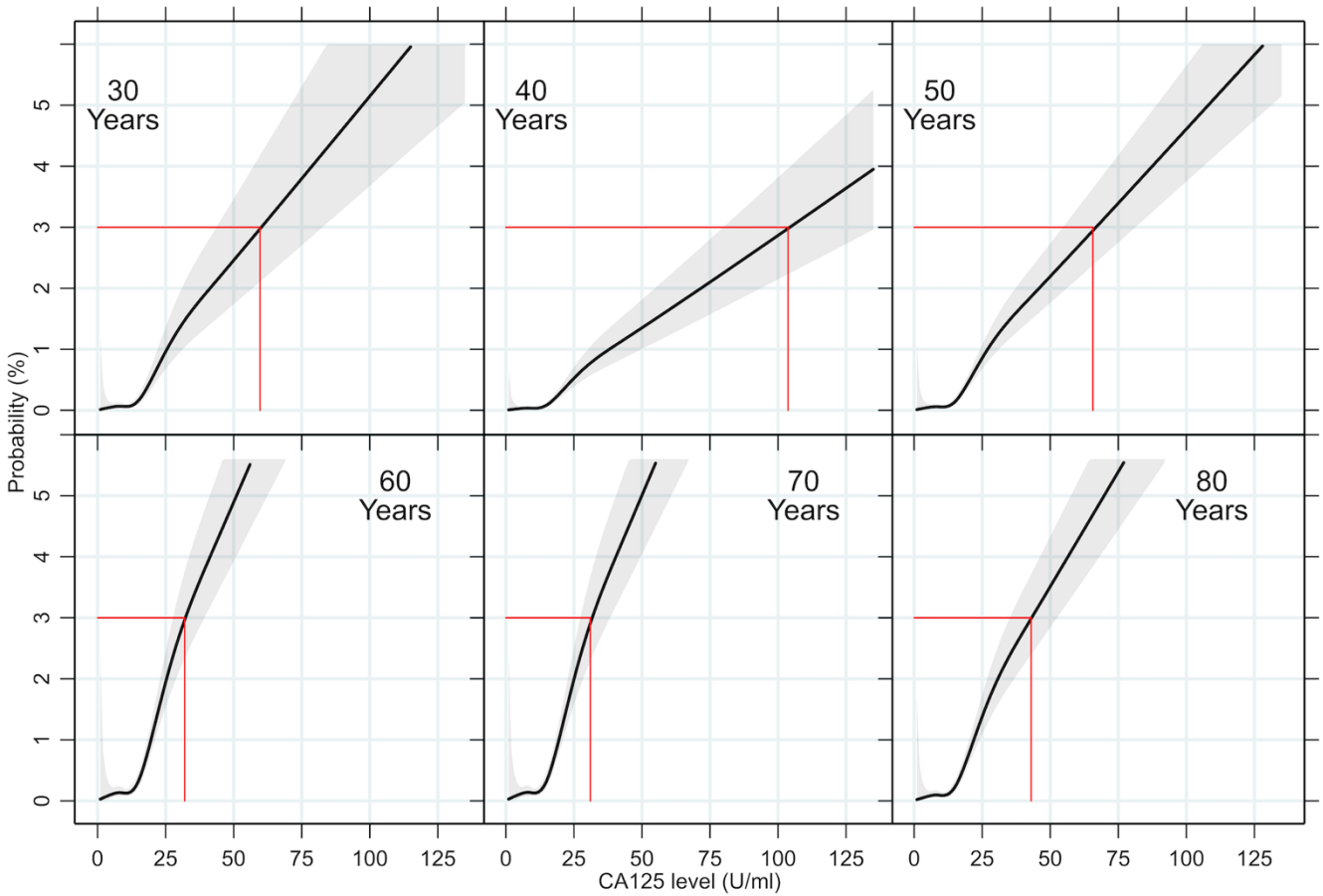


Figure 3.7. Relationship between CA125 level and estimated probability of ovarian cancer for women at age 30, 40, 50, 60, 70 and 80. CA125 levels which correspond to the closest integer probabilities of 3% are indicated in red. 95% confidence intervals are displayed in grey. Prof Gary Abel provided Stata code which was used to lay out this figure.

3.5 Discussion

3.5.1 Summary of key findings

In this cohort study of 50,780 women who had a CA125 test performed in English general practice, 10.1% of those with a CA125 level at or above the conventional cut-off (35 U/ml) were diagnosed with ovarian cancer and 12.3% were diagnosed with a different type of cancer. Just under a third of women aged ≥50 years with a CA125 ≥35 U/ml were diagnosed with some form of cancer. A CA125 level of 53 U/ml equated to an overall ovarian cancer probability of 3%. However, there was marked variation between women of different ages with the 3% probability of ovarian cancer reached at lower CA125 levels in 70-year-old women than in younger or older women. Key findings are summarised against study objectives in **Box 3.1.**

i) **To evaluate the diagnostic accuracy of CA125 for a) ovarian cancer, b) non-ovarian cancers and c) all cancers combined, when used in English primary care.**
   - At the conventional cut-off (≥35 U/ml), CA125 had a:
     - PPV of 10.1% for ovarian cancer and 21.2% for all cancers
     - PPV of 12.3% for non-ovarian cancer (following exclusion of ovarian cancer)
     - Sensitivity of 77% for ovarian cancers, 41.4% for all cancers and 29.1% for non-ovarian cancer
     - High specificity: 93.8% for ovarian cancer, 94.4% for all cancers and 94.4% for non-ovarian cancer

ii) **To explore variations in the diagnostic accuracy of CA125 by patient age**
   - CA125 exhibited superior PPV, sensitivity and specific in women aged ≥50 than <50
   - 15.2% of women ≥50 years with a raised CA125 level had ovarian cancer
   - 20.4% of women ≥50 years with a raised CA125 level who did not have ovarian cancer had another type of cancer

iii) **To explore the association between CA125 level and estimated probability of a) ovarian cancer and b) all cancers combined, in women undergoing CA125 testing in English primary care**
   - The estimated probability of cancer rose with CA125 level, reaching 31.6% for ovarian cancer and 60.6% for all cancers at 500 U/ml
   - The estimated probabilities at each CA125 level (1-1000 U/ml) were calculated and made freely available

iv) **To identify the CA125 level at which a 3% probability of a) ovarian cancer and b) all cancers combined was reached in women undergoing CA125 testing in English primary care**
   - A CA125 level of 53 U/ml equated to a 3% estimated probability of ovarian cancer
   - A CA125 level of 18 U/ml equated to a 3% estimated probability of all cancers

v) **To explore variations in estimated cancer probability by age and determine the CA125 level at which a 3% probability of a) ovarian cancer and b) all cancers combined is reached in different ages**
   - Cancer probabilities varied markedly by age e.g. 3% probability of ovarian cancer was reached at:
     - Age 40: 104 U/ml
     - Age 70: 32 U/ml

Box 3.1. Key study findings against objectives.

3.5.2 Study limitations

I developed a simple model (based on CA125 level) and a multivariable model (based on CA125 level and age) in order to estimate the probability of cancer. Whilst a large cohort of women was selected from a data source which has wide national coverage and is generally representative of the population in England (the CPRD), the models were not validated. Ideally, the models should now undergo external validation in a separate data set to ensure reproducibility and generalisability before they are used in clinical practice.

I employed restricted cubic splines to model the non-linear nature of the relationship between both age and CA125 level and cancer diagnoses. These provide a highly flexible approach to parametrising the fitted relationships. However, there is a large degree of uncertainty at the extremes of age and CA125 level, and so the cancer probabilities for very old and young women and those with very low CA125 levels should be treated with caution and interpreted within the context of the wide confidence intervals.

The results of this study reflect real world use of CA125 in English primary care. International guidelines on CA125 testing vary, and how CA125 is used in primary care in other countries is likely to differ from practice in England.[116] Baseline CA125 levels can be affected by population characteristics, such as ethnicity,[210] which also vary from country to country. Therefore, some caution is needed when translating the study findings to other countries and healthcare systems. If the models developed in this study are to be used in a country with a markedly different healthcare system or population demographics, they should ideally be validated using data from that country.

Previous studies have shown that CA125 has a lower sensitivity for the detection of early stage ovarian cancer than late stage ovarian cancer.[119] Ovarian cancer staging data was contained within my dataset, but I took the decision not to calculate diagnostic accuracy metrics or cancer probabilities by stage. This was because the 12-month follow-up period used in this study provides ample time for cancers which are not immediately referred and diagnosed to progress from early to late stage. Therefore, any stage specific diagnostic accuracy metrics calculated from this study are likely to be misleading.

Given the nature of the data used in this study, it was not possible to ascertain with certainty why CA125 tests were requested, and I chose not to restrict the cohort based on symptom codes, as discussed in **Section 3.3.2**. Some symptoms are more strongly predictive of ovarian cancer than others,[91] therefore the probability of cancer may vary based on the reason for testing e.g. fatigue vs pelvic mass. Other patient factors may also influence the probability of cancer. However, rather than developing a comprehensive diagnostic prediction model, this study sought to explore variation in the probability of cancer based on CA125 level in women of different ages in order to inform clinical practice. In **Chapter 6**, I build on this work to develop a much more comprehensive multivariable prediction model.

3.5.3 Comparison with existing literature

Ovarian cancer

When producing CG122 in 2011, NICE estimated, on the basis of the most relevant data available at the time, that 0.81% of symptomatic women in primary care with a CA125 ≥35 U/ml would have ovarian cancer.[114] Economic modelling and the recommendation for sequential testing with CA125 followed by ultrasound if the CA125 result were abnormal, was predicated on this estimate. The research presented in this chapter indicates that the PPV is more than twelve times higher than estimated by NICE. This finding is consistent with the only other UK report of the PPV of CA125 in primary care, which found that 16 out of 152 women (11%) with a raised CA125 level had ovarian cancer.[156] To estimate the PPV, NICE used diagnostic accuracy data from secondary care studies and, based on primary care research, assumed a prevalence of ovarian cancer in patients presenting with relevant symptoms of 0.23%.[114] However, GPs do not test all women presenting with symptoms such as abdominal pain or bloating for CA125; frequently there is an obvious (non-ovarian) cause for these symptoms. Previous research has shown that women selected for blood tests within primary care have a greater probability of having cancer than matched controls not selected for blood tests.[175] So, it is not surprising that, in this study, the prevalence of ovarian cancer in CA125-tested women was four times higher than the figure used by NICE in their PPV estimates. This highlights the importance of evaluating the performance of tests as they are used in a given setting and population whenever possible, rather than relying on extrapolated data.

The sensitivity of CA125 for ovarian cancer in this study was slightly lower – and the specificity was higher – than in published secondary care studies where testing was performed in women with a pelvic mass prior to surgery.[125] This was expected as tests generally have lower sensitivity and higher specificity in populations with a lower disease prevalence – the spectrum effect (**Chapter 1**).[157] As anticipated, the PPVs for ovarian cancer in my cohort were lower than in secondary care patients with pelvic masses (in whom the prevalence of ovarian cancer is relatively high),[119] and higher than in asymptomatic screening populations (in whom the prevalence of ovarian cancer is relatively low).[152]

This study found that the estimated probability of ovarian cancer for a given CA125 level rose with age to peak in women in their 70's, which mirrors UK age-specific cancer incidence rates.[5] The exception to this trend was very young women; the probability of ovarian cancer at a given CA125 level was higher in women aged 30 than aged 40. This may reflect GP testing practices in very young women (in whom ovarian cancer is extremely rare), with GPs having a strong rationale for requesting a CA125 test, thereby raising the pre-test probability.

Non-ovarian cancer

I believe that the most striking finding from this study was the high incidence of non-ovarian cancer in women with elevated CA125 levels, particularly those aged 50 years or more. This reflects the non-specific nature of ovarian cancer symptoms, and also that CA125 is frequently raised in women with a variety of non-ovarian malignancies.[120] The only comparable primary care study, by Crawford *et al,* reported that 16 out of 152 women (11%) referred from primary care with a raised CA125 were diagnosed with a non-ovarian cancer.[156] Even in asymptomatic screening populations, where symptom overlap does not play a role, a higher incidence of non-ovarian cancers has been noted in women with raised CA125 levels (6.9%) than women with normal CA125 levels (1.6%).[211]

My decision to include non-ovarian cancers in this study was based on reports within the literature that elevations in CA125 can occur in specific types of non-ovarian cancer. While a systematic review of the diagnostic accuracy of CA125 for non-ovarian cancers was beyond the scope of this thesis, some comparison with published literature is warranted. **Table 3.6** compares the sensitivity of CA125 in this study for common CA125-elevating non-ovarian

cancers, with values obtained from the literature. The sensitivity of CA125 for pancreatic and endometrial/uterine cancer are similar between this study and several previous studies, while the sensitivity in this study is higher for lung cancer and lower for breast cancer than previously reported. These differences may be due to variations in case mix and cancer stage between studies.

Table 3.6. Comparison of the sensitivity of CA125 (≥35 U/ml cut-off) for a variety of non-ovarian cancers: current study *vs* the published literature

	Sensitivity (%)	
Cancer type	**Current study**	**Published studies**
Pancreatic	49	45 - 61[126–128]
Lung	47	29 - 30[119,129]
Endometrial/uterine	36	21 - 35[130,131]
Breast	8	14 - 18[119,142]

3.5.4 Clinical relevance of findings

Testing in younger women

39% of CA125 tests were performed in women <50 years of age. However, ovarian cancer is predominantly a disease of older and postmenopausal women. This is reflected in the study findings, as only 18% of ovarian cancers and 11% of invasive ovarian cancers occurred in women <50 years of age. All measures of test performance (save for NPV) were worse in women <50 years than ≥50 years. This was still the case when borderline tumours, which were more common in the younger age group and have less of a propensity to elevate CA125 levels, were excluded. A greater proportion of invasive tumours in the <50 years group were mucinous epithelial and non-epithelial cancers, both of which have less propensity to elevate CA125 than serous epithelial cancer.[119,124] Coupled with lower disease incidence and higher baseline CA125 levels, this likely contributed to poorer test performance in younger women.

The results of regression analysis presented in this chapter indicate that, overall, only 1 in 110 women <50 years with a CA125 of exactly 35 U/ml will have an ovarian cancer and only 1 in 308 will have an invasive ovarian sub-type. Investigating younger women for ovarian cancer when there is a strong clinical suspicion is important, but given the low incidence of ovarian

cancer and relatively poor test performance in women under 50 years, I believe that CA125 tests should be performed and interpreted with caution in this group.

Non-ovarian cancers

The numbers of women in the study cohort with a raised CA125 level who were diagnosed with each particular type of non-ovarian cancer was small, and the test sensitivity for most of these cancers was modest. In isolation, CA125 is unlikely to be a useful test for the detection of individual types of non-ovarian cancer in primary care. However, when considered collectively, the total number of non-ovarian cancers diagnosed in women with raised CA125 levels actually exceeded that of ovarian cancers and, in women ≥50 years of age, a fifth of those with a raised CA125 who did not have ovarian cancer had another cancer type. Given this finding, a high CA125 level in a woman in primary care, especially if aged ≥50 years, should raise a suspicion of non-ovarian cancer. Clinicians should consider these cancers and whether further investigation is needed, particularly if ovarian cancer has been excluded.

Cancer probabilities

The PPV, which is often interpreted as the probability of a disease in patients with a positive test result, was calculated in this study for the conventional CA125 cut-off. Not unexpectedly, the regression analyses revealed that women with very high CA125 values had a very high estimated probability of being diagnosed with cancer. Conversely, those with CA125 levels around the conventional 35 U/ml cut-off had a much lower probability of being diagnosed with cancer than might be expected from the PPV. In this study I have quantified the risk of cancer in individuals with specific CA125 values at specific ages. This should be of more use clinically than the PPV.

The estimated cancer probabilities should allow women and clinicians to interpret their individual CA125 test result and could inform health policy. For example, NICE currently recommends that women with a CA125 ≥35 U/ml, whether 35 U/ml or 350 U/ml, should be referred for an ultrasound scan, while no further investigations for ovarian cancer are advocated in women with levels below the cut-off. This means that some younger women with very low probabilities of ovarian cancer meet referral criteria, while some older women with estimated ovarian cancer probabilities in excess of 3% do not. Instead of using a single

35 U/ml cut-off, the results of this study could be applied to triage women of different ages, selecting those with a high estimated probability of ovarian cancer for expedited referral and investigation. Women with lower probabilities might, after discussion between clinician and patient, be investigated using routine ultrasound, recognising the fact that patients would opt for cancer testing at risk levels as low as 1%.[200] Only a woman's age and CA125 level are required to determine the cancer probability, so this information could readily be incorporated into laboratory IT systems, reported alongside the CA125 level and communicated to patients in clear terms e.g. "1 in 20 women of your age who have the same CA125 level in general practice will have ovarian cancer". The implications of using different thresholds in clinical practice is considered further in **Chapter 6** and potential approaches to integrate CA125 based models within the ovarian cancer diagnostic pathway are considered further in **Chapter 7**.

3.5.5 Conclusions

CA125 is a useful test for ovarian cancer detection in primary care, particularly in women aged 50 years or more. This study has revealed that there is a high incidence of non-ovarian cancers in women with elevated CA125 levels in primary care. Given this, clinicians should consider alternative cancers when confronted with a patient with an elevated CA125, particularly if ovarian cancer has been excluded.

The results of this study will enable patients and clinicians to interpret their individual CA125 result in terms of the probability of cancer at the pertinent CA125 level and age, which should aid informed decision making. The findings will also enable policy makers to provide recommendations for post-CA125 investigations on the basis of the probability of undiagnosed cancer. This could facilitate the expedited investigation and referral of those women most likely to have the disease while providing reassurance for women who are unlikely to have cancer.

Chapter 4. Associations between primary care CA125 test result with test-to-diagnosis interval and stage in ovarian cancer

4.1 Introduction

In the research described in **Chapter 3** I found that, in English primary care, CA125 had a sensitivity of 77% for ovarian cancer at the conventional ≥35 U/ml cut-off. This indicates that, as it is currently used, the test 'misses' 23% of cases. NICE guidelines do not provide specific recommendations regarding the follow-up or further investigation of women with CA125 levels <35 U/ml.[57] Some clinicians have expressed concerns that using CA125 as the sole first line test for ovarian cancer could result in delayed diagnosis and worse outcomes in women with ovarian cancers not associated with high CA125.[212] CA125 levels are less commonly elevated in early stage than late stage disease, so their main concern is that delay in diagnosis and treatment will provide time for early stage (usually curable) cancers to progress to late stage (usually incurable) cancers.

In this exploratory study, I set out to determine whether having a normal (false negative), rather than an abnormal (true positive), pre-diagnostic CA125 test was associated with a longer time between testing and diagnosis. In addition, I sought to examine the association between CA125 test result and stage at diagnosis. As sub-types of ovarian cancer vary both in their propensity to elevate CA125 level and in their aggressiveness, I also wished to compare tumour type between women with normal and abnormal CA125 levels. Through these analyses, I addressed the following thesis objective:

vi. To examine the association of pre-diagnostic primary care CA125 result with time between testing and diagnosis, tumour morphology and disease stage in ovarian cancer.

A paper, based on the work within this chapter, was recently published in the British Journal of General Practice (**Appendix A**):

CA125 test result, test-to-diagnosis interval and stage in ovarian cancer at diagnosis: a cohort study using electronic health records. Garth Funston, Luke TA Mounce, Sarah Price, Brian Rous, Emma J Crosbie, Willie Hamilton, Fiona M. Walter. Br J Gen Pract. 2021; BJGP.2020.0859 (Online First)

4.2 Methods

4.2.1 Study design and participants

This study followed a retrospective cohort design using data from the CPRD, the NCRAS and the Small Area Levels dataset. The study cohort consisted of women, identified from the thesis baseline cohort (**Chapter 2**), who had a diagnosis of ovarian cancer (as recorded within NCRAS data) in the 12 months following their initial CA125 test. Essentially, the cohort consisted of women who had the primary clinical outcome (ovarian cancer) in **Chapter 3**.

4.2.2 Explanatory variables

The principal explanatory variable of interest in this study was the CA125 test result. I also wished to control my analyses for key baseline patient characteristics, which might also be related to the study outcomes. In this section, I describe the explanatory variables and their preparation.

CA125 test result

In this study, I was primarily interested in exploring associations between false negative / true positive CA125 test results and patient outcomes. I therefore used the index CA125 level and applied the conventional ≥35 U/ml cut-off, categorising women into two groups: CA125-normal (<35 U/ml) and CA125-abnormal (≥35 U/ml).

Although this thesis does not focus on the utility of serial CA125 tests, it is possible that retesting might be useful in some circumstances e.g. borderline CA125 levels or persistent symptoms.[55] Some international guidelines recommend retesting in certain circumstances. For example, in Ireland, interval CA125 retesting is recommended in primary care after 6 weeks if an ultrasound, performed due to a high CA125 level, does not show evidence of cancer.[107] I identified women in my cohort who had a repeat CA125 test performed (after the

index date but before the date of diagnosis) and describe these test results within this chapter.

Age

CA125 levels vary with age, and published research indicates that there is an association between older age and longer primary care intervals in ovarian cancer.[213] Given this, I felt that it was important to account for age within my analyses. Age in years, determined during baseline cohort preparation (**Chapter 2**), was identified for each patient included in this study.

Symptom coding

As discussed in **Chapters 2** and **3**, symptoms are not always coded in primary care records prior to a cancer diagnosis, and coding may occur when they are more concerning, severe or persistent. Given this, I postulated that the coding of symptoms might be associated with expedited referral and diagnosis and decided to include coding / lack of coding as a co-variable in this study. A pre-developed list of Read codes, which maps to symptoms included in current NICE guidelines, was provided by Professor Willie Hamilton.[82] From this list, I selected codes for symptoms of ovarian cancer in current NICE guidelines: abdominal / pelvic pain, abdominal distension / bloating, change in bowel habit, fatigue, weight loss, urinary frequency / urgency, loss of appetite, pelvic mass or ascites (**Appendix C**). I used these to identify women with a relevant symptom code in the 30 days before CA125 testing. 30 days was chosen as I was interested in symptoms related to CA125 testing - a month seemed a sufficient period to allow a CA125 test, requested in response to a given coded symptom, to be performed. I generated a two-category variable: “no symptom code recorded” and “symptom code recorded”.

Deprivation

Research has demonstrated an association between higher deprivation and late stage diagnosis in ovarian cancer.[214] I chose to account for deprivation in my analyses. As discussed in **Chapter 2**, the Small Area Levels data within my linked dataset provided a Townsend score for each patient. This score was included as a five-level variable in my analyses.

4.2.3 Outcome variables

I was interested in determining which types of ovarian cancers occurred in women with normal and abnormal index CA125 tests and in examining associations between CA125 test result and a) time between testing and diagnosis and b) stage at diagnosis. In this section, I describe the preparation of tumour morphology and the outcome variables.

Test-diagnostic interval

I calculated the number of days from the index test date to the date of diagnosis for each woman in the cohort. I refer to this period as the 'test-diagnosis interval' throughout this chapter.

Tumour morphology

The preparation of tumour morphology data is described in detail in **Chapter 2** and the morphology of cancers is presented by age group in **Chapter 3**. The greatest difference in behaviour and prognosis in ovarian cancer lies between borderline tumours and invasive tumours, so I first split cancers into these two groups. The distinction between invasive epithelial (usually aggressive) and non-epithelial (frequently indolent) tumours was also important. So, Invasive tumours were further categorised as 'epithelial', 'non-epithelial' or, when ICD-O codes were non-specific, as Invasive Not Otherwise Specified (NOS). This gave four categories:

- Borderline
- Invasive epithelial
- Invasive non-epithelial
- Invasive NOS

Where possible, I also identified details of the histology of invasive epithelial tumours (e.g. clear cell and serous), and describe these within this chapter.

Tumour stage

The determination of cancer stage from NCRAS data is described in detail in **Chapter 2**. Where it is recorded, stage is categorised within NCRAS data as I, II, III or IV. The largest difference in

cancer outcomes in ovarian cancer exists between stage II and III, with 5-year net survival falling from 68% in stage II to 27% in stage III.[16] Due to this, stage I and II ovarian cancer are often regarded as "early stage" and stage III and IV as "late stage" in ovarian cancer research.[91,215,216] I adopted this approach for my analysis.

4.2.4 Statistical analysis

Different methods were required to compare tumour morphology between women with normal and abnormal CA125 test results, to examine the association between CA125 test result and length of test-diagnosis interval, and to examine the association between CA125 test result and stage at diagnosis. I set out each of these approaches separately within this section.

Test-diagnosis interval

A variety of different statistical approaches can be used to examine associations between variables and the length of pre-diagnostic intervals in cancer. Studies have employed linear regression,[217] quantile regression,[218] and survival models,[219,220] amongst others. Accelerated Failure Time (AFT) models are a form of parametric model for survival or 'time to event' analysis. They have previously been used in CPRD diagnostic interval research.[220] In AFT models, covariates act to accelerate or decelerate the time to an event (the 'failure'). Time Ratios (TR's) can be calculated, with a TR >1 indicating that a variable is associated with a prolonged time to an event and a TR <1 indicating that a variable is associated with an earlier event. For example, a TR of 1.5 indicates that a variable is associated with a 50% longer time to event than the comparator. I chose to use AFT models for my test-diagnosis interval analysis, in part because the interpretation of TRs is so simple and intuitive.

I used the *streg* package within Stata to perform AFT model analyses to examine the association between variables and test-diagnosis interval.[221] I first performed univariate analyses for each of the variables: CA125 test result (abnormal or normal), age in years (continuous variable), coded symptom (presence or absence) and Townsend score (five-level). I then constructed a multivariable model which included all of these variables. The Wald test (*testparm* command in Stata) was used to assess the significance of the categorical Townsend score within the model.[222]

As AFT models are parametric, they assume a particular probability distribution in the data. Stata supports exponential, Weibull, log-normal, log-logistic and generalised gamma distributions when using AFT models.[221] I performed analyses applying each of these distributions in turn. I then calculated and compared the AIC of the models. The log-logistic distribution provided the best-fit parameterisation (lowest AIC) and so I took it forward.

I report TRs and associated 95% CI for univariate and multivariate analyses from log-logistic AFT models.

Tumour morphology

First, I described the numbers and proportions of women with each type of ovarian cancer morphology by CA125 group. I used the two-tailed Fischer's exact test to determine whether women with normal and abnormal CA125 levels differed significantly in their tumour morphology. Fischer's exact test is similar to the Chi-square test, but is preferred when the counts in groups or categories are small.[223] First, I ran the test with morphology classified into four categories (borderline, invasive epithelial, invasive non-epithelial, invasive NOS). I then re-ran the test with tumours classified into two groups: borderline and invasive (invasive epithelial, invasive non-epithelial and invasive NOS).

Tumour stage

Not all women with a diagnosis of ovarian cancer had a stage recorded in NCRAS data. Previous cancer studies have used different approaches to handle missing stage data, including treating all cancers with missing staging data as late stage (the 'missing-is-late' approach) or performing a 'complete case' analysis.[224–226] A recent modelling study, which compared these two approaches against a 'gold standard' imputation approach (which utilised a range of variables, including survival time, to impute missing stage), reported that a complete case approach produced less biased results than a missing-is-late approach.[225] I followed a complete case approach, only including women with complete staging data in my analyses of stage at diagnosis. In this subgroup, I used logistic regression analyses to examine associations between CA125 result and stage categorised as 'early' (I-II) or 'late' (III-IV). I first performed univariate analyses for each of CA125 test result, age, presence of a recorded

symptom and Townsend score, then constructed a multivariable model including all these variables.

Given the more indolent course and favourable prognosis of borderline ovarian tumours, I excluded them and performed a sub-analysis for women with invasive ovarian tumours.

In order to assess whether the CA125 test result was associated with missing stage, I ran logistic regression models in which CA125 test result was the independent variable and whether women had staging data 'recorded' *vs* 'missing' formed the binary dependant variable.

I report crude and adjusted odds ratios (OR) with 95% confidence intervals and associated p-values for all analyses.

4.3 Results

4.3.1 Cohort and baseline characteristics

Of the 50,780 women within the baseline cohort, 456 had a diagnosis of ovarian cancer, recorded within NCRAS in the 12 months following the index CA125 test, and were included in this study. 105 women (23%) had a normal index CA125 test and 351 (77%) an abnormal index CA125 test (**Figure 4.1**).

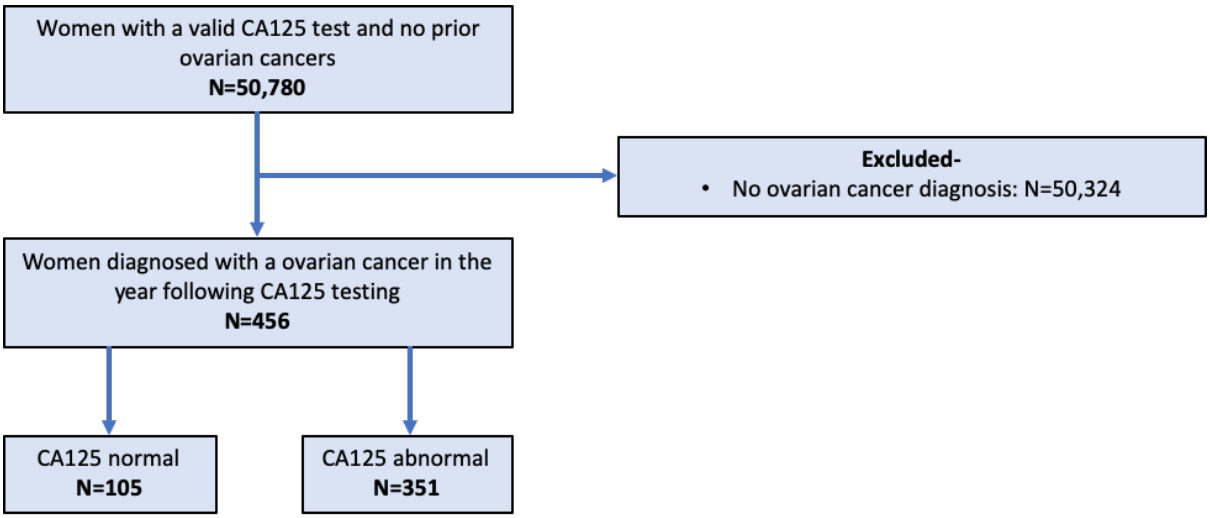


Figure 4.1. Identification of study cohort from baseline cohort.

Mean age was higher in women with an abnormal CA125 result (65 years) than a normal CA125 result (57 years). In addition, a greater proportion of women with an abnormal CA125 result (60.4%) than a normal CA125 result (56.2%) had a coded symptom of possible ovarian cancer (**Table 4.1**). Townsend score was similar between those with normal and abnormal CA125 results.

Table 4.1. Cohort baseline characteristics.

	N	Mean age at diagnosis [range]	Patients with a symptom of possible ovarian cancer recorded pre-testing, N (%)*	Townsend score, N (%)*
Abnormal CA125	351	65 [22-93]	212 (60.4)	Level 1: 80 (22.8) Level 2: 100 (28.5) Level 3: 78 (22.2) Level 4: 61 (17.4) Level 5: 32 (9.1)
Normal CA125	105	57 [18-87]	59 (56.2)	Level 1: 24 (22.9) Level 2: 31 (29.5) Level 3: 25 (23.8) Level 4: 14 (13.3) Level 5: 11 (10.5)
Overall cohort	456	63 [18-93]	271 (59.4)	Level 1: 104 (22.8) Level 2: 131 (28.7) Level 3: 103 (22.6) Level 4: 75 (16.5) Level 5: 43 (9.4)

*Percentage of each group with symptoms and Townsend score

4.3.2 Repeat CA125 testing

41 (9%) women had a repeat CA125 test performed prior to diagnosis. 30 women with an abnormal index CA125 had a repeat test, of which 29 (97%) remained abnormal. 11 women with a normal index CA125 test had a repeat test; 8 (73%) had an increase in their CA125 level, but in only 3 cases (27%) was this increase sufficient to reach the ≥35 U/ml threshold.

4.3.3 Tumour morphology

The types of ovarian cancer which occurred in women in the cohort are summarised by CA125 test result in **Table 4.2**. Borderline tumours accounted for almost half (49%) of tumours in women with normal CA125 results compared to 13% of tumours in women with abnormal CA125 tests. Invasive epithelial cancers were the most common category of tumour in women with abnormal results (81%). Examining the histology of invasive epithelial tumours revealed

that 52% of invasive tumours in the CA125 abnormal group were serous epithelial compared to 30% in the CA125 normal group.

Tumour morphology differed significantly by CA125 test result, both when categorised broadly as 'borderline' and 'invasive' and when categorised into the four groups (borderline, invasive epithelial, invasive non-epithelial and invasive NOS) (p<0.001).

Table 4.2. Tumour morphology by CA125 result.

	Total, n	Borderline, n (%)	Invasive			Overall analysis, p-value
			Epithelial, n (%)	Non-epithelial, n (%)	NOS, n (%)	
Abnormal CA125	351	47 (13)	284[a] (81)	4 (1)	16 (5)	<0.001
Normal CA125	105	51 (49)	39[b] (37)	9 (9)	6 (6)	

NOS = Not otherwise specified i.e. could not be classified as epithelial or non-epithelial based on the information within the cancer registry. P-value derived from Fisher's exact test for independence.

[a] 158 serous, 16 endometrioid, 14 mucinous, 14 clear cell, 13 other epithelial and 69 epithelial cancers of unknown morphology.

[b] 16 serous, 4 endometrioid, 8 mucinous, 3 clear cell, 4 other epithelial and 4 epithelial cancers of unknown morphology.

4.3.4 Test-to-diagnosis interval

Histograms of the test-diagnosis intervals in women with normal and abnormal CA125 tests results are shown in **Figure 4.2**. The median test-to-diagnosis interval in the cohort was 42 days (interquartile range [IQR]: 25-62 days). In women with abnormal CA125 results it was 35 days (IQR: 21-53 days) compared to 64 days (IQR: 42-127 days) in women with normal CA125 results. 2% of women with abnormal CA125 results had a test-diagnosis interval of more than 6 months compared to 15% of women with normal CA125 test results.

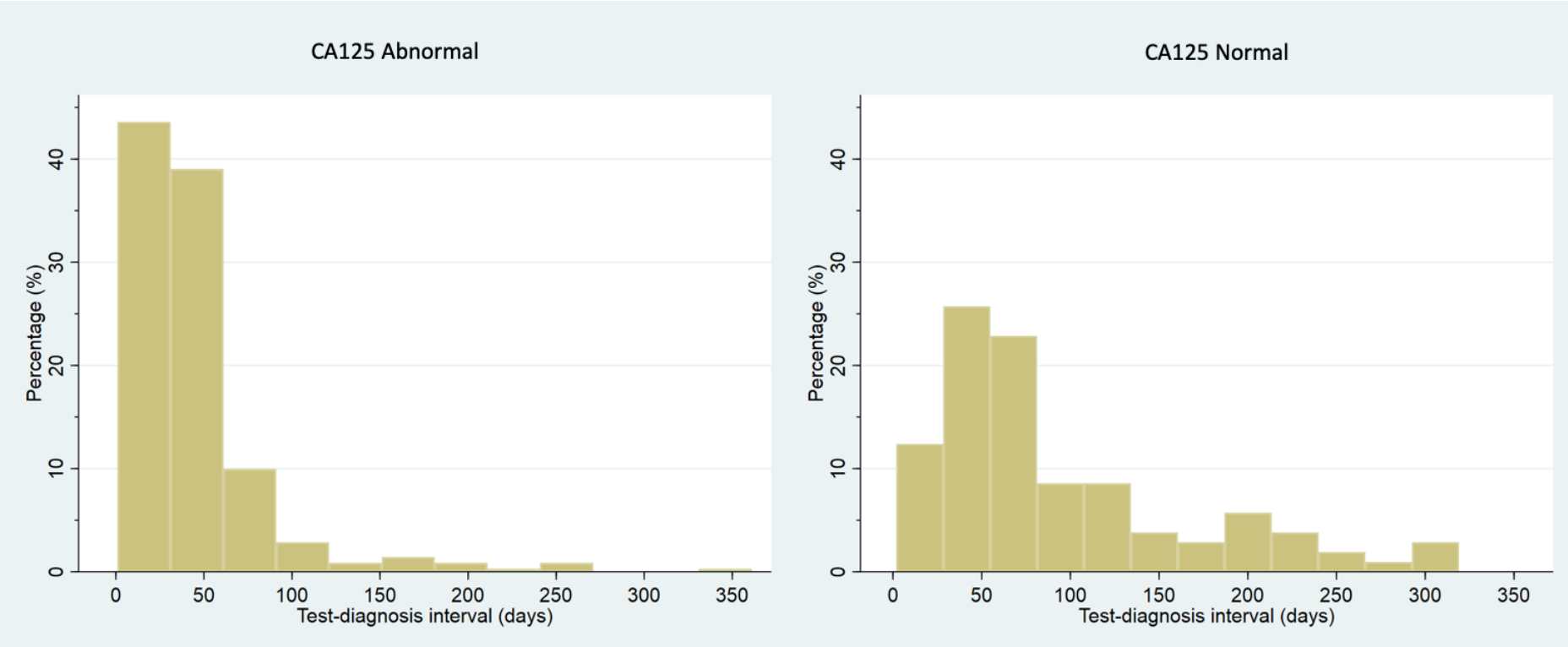


Figure 4.2. Histograms showing the distribution of test-diagnosis intervals in women with normal and abnormal CA125 tests.

In a univariate AFT analysis, having a normal CA125 result was significantly associated with a longer test-diagnosis interval (**Table 4.3**). The TR of 2.0 (95% CI:1.7-2.4, p<0.001) indicates that the test-to-diagnosis interval was twice as long in women with normal than abnormal CA125 test results. In the multivariable analysis, the adjusted TR for CA125 test result remained unaltered. On univariate analysis, higher age was significantly associated with a slightly shorter test-diagnosis interval (TR 0.99, 95% CI: 0.99-0.99, p=0.004), but this association was not present in the multivariable analysis (TR: 1, 95% CI: 1.0-1.0, p<0.16).

Table 4.3. Crude and adjusted associations between CA125 test result, age, presence / absence of a coded symptom of possible ovarian cancer and Townsend score with Test-to-diagnosis intervals.

	Unadjusted		Adjusted	
	TR (95% CI)	p value	TR (95% CI)	p value
Abnormal CA125	Reference	-	-	-
Normal CA125	2.0 (1.7-2.4)	<0.001	2.0 (1.6-2.4)	<0.001
Age (years)	0.99 (0.99-0.99)	0.004	1.0 (1.0-1.0)	0.16
No symptom record	Reference	-	-	-
Symptom record	0.95 (0.80-1.1)	0.50	0.94 (0.81-1.1)	0.46
Townsend score	-	0.75[a]	-	0.77[a]
Townsend 1	Reference	-	-	-
Townsend 2	1.0 (0.83-1.3)	0.70	1.0 (0.82-1.3)	0.87
Townsend 3	1.1 (0.89-1.4)	0.33	1.1 (0.86-1.3)	0.51
Townsend 4	0.97 (0.75-1.3)	0.80	1.0 (0.77-1.2)	0.85
Townsend 5	1.1 (0.82-1.5)	0.50	1.0 (0.77-1.4)	0.81

[a] Derived using the Wald test.

4.3.5 Stage at diagnosis

Staging information was missing for 75 (16%) women. A greater proportion of women in the normal CA125 group had missing stage (n=28, 27%) than in the abnormal group (n=47, 13%). In the restricted cohort of patients with a recorded stage, baseline characteristics were very

similar to the main ovarian cancer cohort (**Appendix F**). In this group (n=381), 172 (45%) had early stage and 209 (55%) had late stage disease. 35% (n=66) of those with an abnormal CA125 result were diagnosed with early stage disease compared to 86% (n=198) of those with a normal CA125 result.

The results of the univariate and multivariable logistic regression analyses, performed to examine the association of CA125 test result and early stage diagnosis, are shown in **Table 4.4**. On univariate analysis, the odds of early stage diagnosis were 11.2 times greater in women with normal than abnormal CA125 test results (95% CI: 5.7-22.1, p<0.001). Increasing age and the presence of a recorded symptom were inversely associated with early stage diagnosis on univariate and multivariable analyses. In the multivariable model, the odds of early stage diagnosis were 12.2 times higher in women with normal than abnormal CA125 test results (95% CI: 5.8-25.15, p<0.001).

In a sub-analysis, conducted after excluding borderline tumours (**Appendix F**), the odds of early stage diagnosis were still 9.1 times greater in women with normal than in women with abnormal CA125 tests (95% CI:4.0-19.8, p<0.001).

Table 4.4. The association between CA125 test result, age and presence / absence of a recorded symptom with early stage (I-II) diagnosis.

	Unadjusted		Adjusted	
	OR (95% CI)	**p value**	**OR (95% CI)**	**p value**
Abnormal CA125	Reference	-	Reference	-
Normal CA125	11.2 (5.7-22.1)	<0.001	12.2 (5.8-25.5)	<0.001
Age (years)	0.95 (0.93-0.96)	<0.001	0.94 (0.92-0.96)	<0.001
No symptom record	Reference	-	-	-
Symptom record	0.51 (0.33-0.77)	0.001	0.35 (0.21-0.59)	<0.001
Townsend score	-	0.4	-	0.9
Townsend 1	Reference	-	Reference	-
Townsend 2	1.8 (1-3.2)	0.05	1.4 (0.7-2.8)	0.3
Townsend 3	1.4 (0.8-2.6)	0.2	1.1 (0.5-2.2)	0.9
Townsend 4	1.6 (0.8-3.0)	0.2	1.2 (0.6-2.6)	0.6
Townsend 5	1.8 (0.8-3.9)	0.2	1.2 (0.4-3)	0.8

* Derived using the Wald test.

In a multivariable model, adjusted for age, symptom coding and Townsend score, the odds of having missing stage were 2.6 times higher in CA125 normal than CA125 abnormal women (95% CI: 1.5-4.6, p=0.001). No statistically significant association was identified when borderline tumours were excluded from this analysis (adjusted OR: 1.7, 95% CI: 0.8-3.8, p=0.8).

4.4 Discussion

Women with normal CA125 results in primary care prior to ovarian cancer diagnosis took twice as long to be diagnosed following testing than those with abnormal results. Despite this, in women for whom staging data was available, 86% of those with normal CA125 results were diagnosed at an early stage compared to only 35% of those with abnormal CA125 results. In addition, indolent borderline ovarian tumours were more common, and aggressive invasive epithelial cancers less common, in women with normal than abnormal CA125 results. Key findings are summarised against the study objective in **Box 4.1**.

vi) To examine the association of pre-diagnostic primary care CA125 result with time between testing and diagnosis, tumour morphology and disease stage in ovarian cancer.

- Median test-diagnosis interval was:
  - 34 days in women with an abnormal CA125
  - 64 days in women with a normal CA125
- Having a normal CA125 test result, rather than an abnormal CA125 test result, was associated with a doubling of the test-diagnosis interval

- Tumour morphology differed significantly between women with normal and abnormal CA125 test results
- Indolent borderline tumours were the most common tumour type in women with normal CA125 levels (49%) while invasive epithelial tumours were the most common tumour type in women with abnormal CA125 levels (81%)

- Having a normal CA125 result was associated with early stage diagnosis. In women with a recorded stage, the odds of early stage diagnosis were 12 times higher in women with a normal rather than an abnormal CA125 test result

Box 4.1. Key results against thesis objective vi.

4.4.1 Limitations

Patients with severe disease, who often have more severe symptoms, frequently experience expedited diagnoses when compared to those with less severe disease. This observation is often referred to as the 'sick quick' phenomenon.[76] As CA125 is also more likely to be elevated in women with more severe disease, this may act as a confounder. I have adjusted the analyses for the presence/absence of relevant coded symptoms, as symptoms may be more likely to be coded (rather than mentioned in free text) if they are more severe.[187] However, I am unlikely to have been able to adjust fully for severity of presentation.

A normal CA125 result was significantly associated with missing stage. This is to be expected, as stage is less frequently recorded in the cancer registry for borderline tumours, which are more common in women with normal CA125s. It is reassuring that when borderline tumours were excluded no significant association between CA125 result and missing stage was identified and that a normal CA125 result was still strongly associated with early stage diagnosis. While I have no reason to suspect that study findings would differ substantially if staging data were available for all patients, the magnitude of the association between CA125 result and stage should be interpreted with caution.

The analysis of CA125 test and stage was exploratory. The observational nature of this study meant that, although I was able to report that women with normal CA125 levels were usually diagnosed at an early stage, I could not determine to what extent women with normal CA125 tests experienced disease progression between having the CA125 test and being diagnosed. Without a controlled trial, or a further evaluation following a change in practice, it is not possible to determine whether shortening the test-diagnosis interval in CA125 normal women would result in earlier stage diagnosis in this group.

4.4.2 Comparison with existing literature

Previous studies have identified an association between false negative results in primary care for several other types of test and longer healthcare intervals. In one study, patients with a normal chest X-ray prior to lung cancer diagnosis experienced longer primary care intervals than those with an abnormal chest X-ray.[227] While in another study, patients with a false negative rheumatoid factor in primary care prior to a rheumatoid arthritis diagnosis took longer to be referred to a specialist.[228] Studies in other healthcare settings also indicate that false negative test results may delay diagnosis and treatment. For example, false negative breast and cervical cancer screening tests have been implicated in diagnostic delay and an association between false negative fine needle aspiration and delayed treatment has also been reported.[229,230] It is possible that negative results could provide 'false reassurance' to patients, thereby delaying their re-presentation if symptoms persist, or provide false reassurance to healthcare professionals, prompting them to seek alternative diagnoses and delaying referral. However, the mechanisms by which delays occur in the presence of false

negative test results are not fully understood and I did not identify any studies exploring mechanism of delay due to false negative CA125 results.

Few studies have investigated the relationship between false negative test results and outcomes in symptomatic patients. However, one study did find that patients with false negative fine-needle aspiration before thyroid cancer diagnosis had worse outcomes (greater vascular and capsular invasion, more frequent persistent disease post treatment).[230] Although women with normal CA125 results in my study took markedly longer to be diagnosed than those with abnormal CA125 results, they were generally diagnosed at an earlier stage. This finding could be due to differences in tumour type between the groups; borderline tumours were four times as common in women with normal CA125 (in whom they were the most common type of tumour) than in women with abnormal CA125 test results. Borderline tumours are known to cause less frequent elevations in CA125 than their invasive counterparts.[204] Borderline tumours tend to grow slowly and spread late, with 80% diagnosed at an early stage.[204] By contrast, invasive epithelial tumours, which typically have an insidious onset and poor survival, were much less common in women with normal than abnormal CA125 test results. Even within the invasive tumour category, serous epithelial tumours, which are particularly aggressive and more frequently diagnosed at a late stage than other invasive tumour types,[119] were less common in women with normal than abnormal CA125 test results. So, while the tumours in women with normal CA125 tests had more time to progress during the test-diagnosis interval, they were predominantly indolent and the majority did not advance from early to late stage during this time period.

4.4.3 Implications for research and practice

In the overall study cohort the median test-diagnosis interval was 14 days longer than the 'Faster Diagnosis Standard', introduced in England in April 2020, which recommends that patients should receive a diagnosis of cancer within 28 days of being referred to a specialist by their GP.[68] In women with abnormal CA125 results, the median test-diagnosis interval was only 7 days longer than this standard. While women with normal CA125 levels did take markedly longer to be diagnosed, it is somewhat reassuring that the majority of these women had indolent tumour types which were predominantly diagnosed at an early stage.

A significant amount of research effort in countries around the world is directed at developing alternative biomarkers to CA125 which are more sensitive for ovarian cancer.[216,231] Given that most women with false negative CA125 test results were diagnosed at an early stage, this study raises the question of how much impact a more sensitive first line testing strategy could realistically have on stage at diagnosis and survival. A more sensitive first line testing strategy in primary care might be expected to reduce the test-to-diagnosis interval by a median of around a month (i.e. the median difference between the CA125 normal and abnormal CA125 groups). However, a much greater proportion of women with normal CA125 levels had extreme diagnostic intervals e.g. >6 months; therefore, reducing the numbers of false negative results could result in a dramatic reduction in the test-diagnosis interval for some women. Although women with normal CA125 results often had more indolent forms of ovarian cancer, 15% of invasive cancers, and 12% of the HGS sub-type (the most aggressive form) occurred in this group. We cannot establish from this study what proportion of women with a normal CA125 result experienced disease progression during the prolonged test-diagnostic interval. However, even a progression within the early stage group (e.g. from stage IA to IC or from I to II) has prognostic implications, particularly for aggressive invasive ovarian cancers. Although there is often a focus on achieving a 'stage shift' in early detection research, earlier diagnosis of ovarian cancer has the potential to reduce morbidity and mortality, even if a stage shift is not achieved, by detecting lower volume (less 'bulky') disease which is more amenable to complete surgical resection.[55] As discussed in **Chapter 1**, complete cytoreduction (the removal of all visible tumour) is recognised as a key prognostic factor in ovarian cancer.[70,72]

Prompt diagnosis also has other benefits. For example, delay in cancer diagnosis is associated with psychological distress and perceived delays can damage doctor-patient relationships.[232,233] So, while this study demonstrates that large proportions of women do not end up with late stage disease as a result of false negative CA125 results (which should provide some reassurance for those using and being tested for CA125), there is potential to improve patient psychosocial outcomes by reducing the number of false negative results. Ultimately, prospective studies, or further retrospective database studies, would be needed to assess the impact of any novel first line testing approach, including the use of alternative thresholds for referral derived from the models presented in **Chapter 2** and **6**.

Current NICE guidelines provide no specific recommendations on follow-up in women with normal CA125 levels and persistent symptoms. However, safety netting and retesting with CA125 if symptoms persist or worsen is a potential option which warrants further evaluation. In this study, only a small proportion (9%) of women with normal initial CA125 results had a repeat test. There was an increase in levels in 73% of cases, but in only 27% of cases was this increase sufficient to reach the 35 U/ml threshold. This supports the idea put forward by some researchers that if a women has a rising CA125 level (even if it is below the 35 U/ml threshold) this should prompt further investigation.[55] The Risk of Ovarian Cancer (ROCA) algorithm, used within the UKCTOCS trial, was developed for this reason and relies on changes from a woman's baseline CA125 level to determine when further investigation is needed.[234] Alternatively, persistent symptoms in those with a normal initial CA125 test could prompt a different modality of testing e.g. ultrasound. Any potential follow-up strategy must take account of the burden of testing on those without ovarian cancer, given that only one in every 451 women with a normal index CA125 test in primary care will have the disease (**Chapter 3**).

4.5 Conclusion

Women with false negative CA125 test results in primary care experienced markedly longer intervals between testing and ovarian cancer diagnosis than women with true positive CA125 results. However, they more frequently had indolent forms of the disease and the majority were diagnosed at an early stage. It is reassuring that a large majority of women with false negative CA125 results did not progress from early to late stage disease as a result of a prolonged test-diagnosis interval. Nevertheless, expediting diagnosis in this group of women still has the potential to improve patient outcomes.

Chapter 5. Symptom based tools for ovarian cancer detection: a systematic review

5.1 Introduction

In **Chapter 1,** I introduced the concept of diagnostic prediction models and highlighted several which are currently available for use in primary care. When planning my doctoral research, I postulated that a diagnostic prediction model, incorporating CA125 levels alongside other variables associated with ovarian cancer, might outperform CA125 alone in the detection of the disease. A key objective of my doctoral research was to develop such a prediction model. Prior to developing it, I wanted to perform a review of the literature in order to identify existing ovarian cancer detection tools and ensure that my planned prediction model (or one very similar to it) had not already been developed. I also wished to summarise the diagnostic performance of existing ovarian cancer detection tools to enable comparison with my model. Determining which variables to include within a prediction model is one of the most challenging aspects of model development. I saw a literature review as a key first step in identifying variables from existing tools which might be relevant for inclusion in my model.

The model I planned to develop was predicated on CA125 i.e. it would be developed in women undergoing CA125 testing and would incorporate CA125 level. So, my initial intention was to perform a systematic review of tools which included CA125 as a variable. However, preliminary searches revealed that most of these tools combined CA125 with specialist biomarkers and/or imaging tests e.g. ultrasound. These specialist tests are not available within UK primary care and the results of imaging tests are not recorded in the CPRD, so I judged that a review focussed on CA125-based models would have been of very limited utility in helping to identify candidate variables for my model.

Instead of performing a systematic review of CA125 based tools, I chose to perform a systematic review of symptom-based tools for ovarian cancer. Symptoms are the trigger for performing CA125 tests in primary care and are predictive of the presence of the disease in

their own right. However, as discussed in **Chapter 1**, not all symptoms are equally predictive. Given this, I saw symptoms as a key variable type for the development of my prediction model. Preliminary searches in MEDLINE revealed that a number of symptom-based tools for ovarian cancer had been developed and that some incorporated other types of variable e.g. cancer risk factors and test results. I determined that a systematic review of these tools would allow me to ascertain whether a prediction model combining CA125 with symptoms and other variables had previously been developed, and would also provide insight into potential candidate variables for my model.

In this chapter I describe the steps I took to identify relevant symptom-based tool studies, provide an appraisal of study quality, summarise the variables included within tools and provide a comparison of their diagnostic accuracy. The thesis aim addressed by this chapter is:

vii. To perform a systematic review to identify published symptom predicated tools for ovarian cancer detection and to compare these tools in terms of a) included variables and b) diagnostic accuracy.

A paper, based on the work presented within this chapter, was recently published in Cancers (**Appendix A**):

Identifying ovarian cancer in symptomatic women: a systematic review of clinical tools. Garth Funston, Victoria Hardy, Gary Abel, Emma J. Crosbie, Jon Emery, Willie Hamilton, Fiona M. Walter. Cancers. 2020; 8:3686

5.2 Methods

5.2.1 Previous studies

Prior to commencing this study, I reviewed several databases of systematic reviews to ensure that a similar review had not been published or was underway. I did not identify any published systematic reviews covering symptom-based tools for ovarian cancer detection. A Cochrane protocol for a series of reviews, designed to evaluate the diagnostic accuracy of tests and symptoms alone and in combination, was published in 2015.[235] It is good research practice to

avoid duplicating systematic reviews (unless performing an update after a reasonable period), as this wastes time, effort and research funding. This contributed to my decision not to perform and publish a formal systematic review on the diagnostic accuracy of CA125 at the outset of my doctoral research. However, as of April 2021 (more than five years after the publication of the Cochrane protocol) none of the Cochrane reviews have been published.

I considered a literature review an essential step to inform the development of a prediction model during my PhD research and so chose to proceed with this systematic review. This review differs from the Cochrane review series in that it only includes tools which incorporate symptoms. In addition, while Cochrane diagnostic accuracy studies impose very stringent quality criteria during study selection, I did not exclude any studies on the basis of quality as I wished to identify all relevant published tools and identify which variables they included in order to inform my model development. So, while there may be a degree of overlap between this study and any reviews which may arise from the Cochrane protocol, I believe the results and their interpretation are likely to differ.

5.2.2 Review registration and reporting

Prior to conducting this review, I developed a study protocol and registered it with PROSPERO [CRD42020149879].

I report the review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines and include a PROSPERO checklist in **Appendix G**.

5.2.3 Search strategy

I developed a search strategy in MEDLINE to include terms for ovarian cancer, symptoms and prediction / diagnostic models / tools. Relevant Medical Subject Heading (MeSH) terms were also identified and included. Prior to performing the definitive search, I piloted the strategy to ensure that it identified several key papers.[54,91,215] The final MEDLINE search strategy is included in **Appendix H**. I converted this strategy for use in Cochrane CENTRAL and EMBASE and searched all three databases from 1st January 2000 – 3rd March 2020. The start date was chosen to predate key ovarian cancer symptom studies.[91,92] I considered it unlikely that symptom-based ovarian cancer detection tools would have been developed prior to the

publication of these symptom studies. No language restrictions were applied. I did not place any restrictions on the methodological design of studies, so both case-control and cohort studies were eligible for inclusion. In addition, no restrictions were placed on study setting, so studies conducted in the general population or the clinical setting were eligible for inclusion. As well as identifying studies through database searches, I screened the reference lists of included papers to identify any additional relevant publications.

5.2.4 Eligibility criteria

Studies were included if they a) described the development and or evaluation of a multivariable tool designed to identify patients with undiagnosed ovarian cancer, and b) provided the sensitivity and specificity of the tool or gave sufficient data to allow these metrics to be calculated. For the purposes of this review, I defined a multivariable tool as a combination of three or more variables used to detect, or predict the risk of, undiagnosed ovarian cancer. During my initial pilot searches I identified tools which took the form of diagnostic prediction models (a "mathematical equation that relates multiple predictors for a particular individual to the probability of or risk for the presence… of a particular outcome").[162] I also identified tools which did not meet the above definition of a prediction model but met the broader definition of a decision / prediction rule (a decision making tool consisting of at least three variables which provides the probability of an outcome or suggests a course of action).[236] I believed that both tools could help me identify candidate variables for my model so included both types of tool in this review. My broad definition of a multivariable tool encompasses formal multivariable diagnostic prediction models and clinical prediction rules.[162,236] I considered variable 'checklists', in which only one variable in the list needs to be present for a positive tool result, to be a form of multivariable tool. As the focus of this review was on symptom-based tools, the tool under investigation had to include one or more symptoms for a study to be eligible. No other restrictions were placed on the types of variable that could be included.

Studies on tools intended to estimate *future* risk of developing cancer rather than the *current* risk of having an undiagnosed cancer were excluded, as were studies on tools that solely provide an indication of the risk of ovarian cancer relapse or recurrence. I excluded studies in which all participants had a known pelvic mass, as this represents a highly selected high-risk

population. I excluded studies undertaken solely in paediatric (<18 years) populations, as I planned to develop my model in women ≥18 years of age who generally develop different types of ovarian tumour. Non-primary research studies e.g. review articles, and conference abstracts were also excluded.

5.2.5 Study selection

I exported the results from database searches into Endnote then added them to Rayyan;[237] a software designed to facilitate duplicate removal and co-ordinate abstract screening by multiple reviewers. I removed duplicates from the searches using Rayyan. Following this, Victoria Hardy (the second reviewer on the study [VH]) and I independently screened the titles and abstracts of identified documents against an eligibility criteria checklist. Potentially eligible papers identified at the screening stage were obtained and then the full texts independently examined by me and VH against the eligibility criteria. We discussed any disagreements and resolved them by consensus.

5.2.6 Data extraction

I developed an Excel template for data extraction. Using this, I extracted information from each paper on: study characteristics (year of publication, location); study design (methodology, population, data source, outcome definition); tools (variables, tool development methods); and tool performance metrics (sensitivity, specificity, other diagnostic metrics). Where a study evaluated multiple tools, I extracted data relating to each tool separately. My extracted information was then checked against the full text papers by VH to ensure accuracy.

5.2.7 Data synthesis

When synthesising data, I paid particular attention to three key study and tool characteristics. First, the source of participant recruitment. For example, whether controls were recruited from the general population or from the healthcare setting (primary / secondary care). Symptoms may be more common in clinical controls (e.g. those attending gynaecology clinics) than in controls recruited from the general population, which could influence estimates of tool diagnostic performance.[157] Second, whether the measures of tool accuracy were obtained directly from the patient sample in which the tool was developed (apparent

performance), by applying internal validation methods such as cross-validation (internal validation) or from an evaluation of the tool in a distinct patient sample (external validation).[162] Tools usually exhibit poorer diagnostic performance in external validation studies than when evaluated in the original development sample, and external validation of tools is generally recommended before they are used in clinical practice.[162] Third, I considered the type of tool being evaluated. Whether a tool is a simple symptom checklist, a checklist of symptoms and other variables (augmented symptom checklist) or a prediction model, is likely to affect its clinical utility.

I tabulated and compared the variables included in each tool. I compared tools in terms of their sensitivity and specificity. For diagnostic prediction models, I compared AUC. AUC is a measure of model discrimination i.e. the ability of a model to identify those with a condition from those without a condition. I also compared model calibration (agreement between estimated and observed outcomes) of models where relevant calibration metrics were provided in papers. When planning this study, I hoped to perform a meta-analysis of diagnostic accuracy metrics. However, there was marked heterogeneity in included studies in terms of their designs, populations, variable definitions, outcome definitions and use of different thresholds. Multiple studies also failed to report numbers of patients with true positive / true negative / false positive / false negative results. Ultimately, I decided that a meta-analysis would not be informative. Instead, I summarised performance characteristics in tabular form and used a narrative synthesis approach.

5.2.8 Risk of bias assessment

I identified two possible tools which could be used to assess the risk of bias (ROB) of included studies: the Prediction model Risk Of Bias Assessment Tool (PROBAST),[238] and the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.[184] There is overlap between the tools, but PROBAST is specifically designed to assess ROB in prediction modelling studies whereas QUADAS-2 is designed to assess diagnostic accuracy studies. Each of these tools was potentially appropriate for assessing ROB in this review, so I piloted them on several studies. As many of the studies in this review evaluated decision rules rather than true prediction models, I found that multiple sections of PROBAST were of limited relevance to these studies. In addition, it failed to assess potential sources of bias in the assessment of tool diagnostic

accuracy in depth. By contrast, all sections of QUADAS-2 were of relevance to the review studies, but it did not fully consider the ROB from statistical approaches used to develop prediction models. Neither tool was ideally suited to assess for ROB in all included studies, given the diverse methods used. However, I chose to use QUADAS-2 in this review as the reported accuracy of tools, rather than the particular statistical approaches used to derive prediction models, was of greater interest.

QUADAS-2 includes signalling questions (intended to identify areas of potential bias or concern over study applicability) covering four domains: 1) patient selection, 2) index test(s), 3) reference standard and 4) flow and timing. Each domain is rated as of "high", "low" or "unclear" (where insufficient information is provided) ROB. In domains 1-3 papers are also rated on their applicability (how well they match the review question), as of "high", "low", or "unclear" concern. VH and I each independently assessed the included studies using QUADAS-2. Ratings were compared and disagreements resolved by consensus.

5.3 Results

5.3.1 Study selection

I identified 2,331 records through database searches, of which 708 were duplicates. On examination of reference lists, I identified two additional relevant documents. A total of 1,625 titles and abstracts were screened and 35 full text papers were obtained and examined against eligibility criteria. 16 studies met these criteria and were included in the review (**Figure 5.1**).

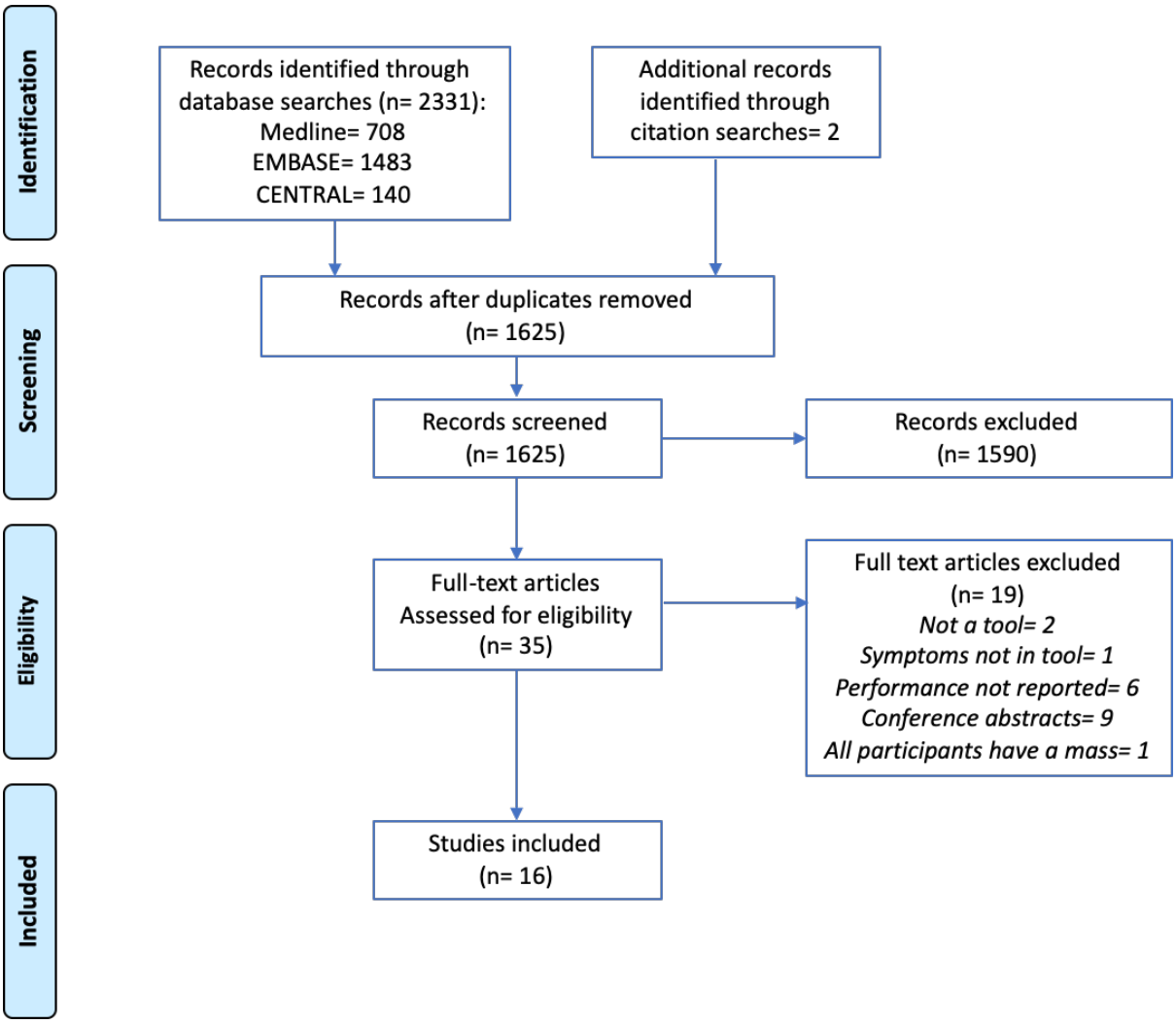


Figure 5.1. PRISMA flow diagram illustrating the study selection process.

5.3.2 Study characteristics

Table 5.1 summarises key characteristics of included studies including setting, design, objectives, outcome and participants. Specific exclusion criteria are provided in **Appendix H**. Three studies were population-based (i.e. recruited participants at the population level),[239–241] five were primary care based,[54,91,242–244] four were entirely hospital-based,[245–248] and four were hospital-based but also recruited controls from screening trials rather than the clinical setting.[215,249–251] All population and hospital-based studies were of case-control design. Two of the hospital based case-control studies included a proportion of controls with benign ovarian pathology.[245,247] Three of the five primary care studies were of cohort design,[54,242,243] and the remaining two of case-control design.[91,244]

The number of women in each study varied markedly, with between 75 – 1,908,467 participants and 24 – 1,885 patients with ovarian cancer per study. A range of different data-sources were used by studies to collect variable information, including surveys or patient interviews (n=11), routinely collected primary care databases (n=6), and blood samples (n=4). All studies used ovarian cancer as an outcome, but how this was defined varied, with some only including invasive cancers or invasive epithelial cancers,[239–241,245–248] and others including both invasive and borderline epithelial tumours or all types of ovarian cancer.[54,91,215,242–244,249–251] One study included ovarian cancer alongside other common cancers in a composite outcome, but did provide tool performance characteristics for each individual cancer.[242] Seven studies developed entirely new tools,[54,91,215,239,242,244,251], six modified existing tools,[245–250] and eight externally validated existing tools.[240,241,243,245–248,251]

Table 5.1. Study characteristics.

Author, date, country	Design		Objective			Primary outcome	Candidate variable data sources	Participants	Study sample
	Case-control	Cohort	Develop new tool	Modify existing tool	Externally validate existing tool				
Population based									
Lurie, 2009, USA	●		●			Primary invasive ovarian carcinoma	In person patient interviews using a structured survey	*Cases:* Women aged 19-88 years, histologically confirmed primary invasive ovarian carcinoma (1993 - 2007) *Controls:* Aged ≥18 years, Hawaii resident ≥1 year, randomly selected from statutory state survey. Frequency-matched to cases (1:1) by age, ethnicity, interview time	*Cases:* 432 *Controls:* 491
Rossing, 2010, USA	●				●	Primary invasive epithelial OC[a]	In person interviews	*Cases:* Resident in western Washington State, aged 35–74 years, diagnosed with a primary invasive epithelial ovarian tumour (Jan 2002 - Dec 2005) *Controls:* Selected by random digit dialling with stratified sampling in 5-year age categories, 1-year calendar intervals, and two (urban vs suburban or rural) county strata	*Cases*: 594 *Control*s: 1,313

Jordan, 2010, Australia	●				●	Invasive epithelial OC	Patient survey	*Cases:* Aged 20-79 years with suspected OC subsequently diagnosed with invasive epithelial OC (Jan 2002 - Jun 2005) *Controls:* Frequency-matched based on age (5-year groups) and state of residence identified from electoral roll[252]	*Cases:* 1215 *Controls:* 1,456
Primary care population									
Hamilton, 2009, England	●		●			Primary OC, including borderline	Researcher coded GP records	*Cases:* Aged ≥40 years with primary OC diagnosed between 2000-2007 *Controls:* Matched on age, sex and GP practice	*Cases:* 212 *Control:* 1,060
Hippisley-Cox, 2012, England and Wales		●	●			OC (NOS)	QResearch database[180]	Aged 30-84 years, registered with GP practices between 1 Jan 2000 - 30 Sept 2010	*Development (2/3)*- 1,158,723 women with 976 OCs *Validation (1/3)*- 608,862 women with 538 OCs
Hippisley-Cox, 2013, England and Wales		●	●			OC (NOS) and 10 other cancers	QResearch database[180]	Aged 25–89 years, registered with GP practices between 1 Jan 2000 - 1 Apr 2012	*Development (2/3)*- 1,240,864 women with 1,279 OCs *Validation (1/3)*- 667,603 women with 606 OCs

Grewal, 2013, England	●		●			Primary OC, including borderline	Researcher coded GP records	*Cases:* Aged ≥40 years with primary OC diagnosed between 2000-2007 *Controls:* Matched on age, sex and GP practice	*Cases:* 212 *Control:* 1,060
Collins, 2013, UK		●			●	OC (NOS)	THIN database[253]	Women 30-84 years registered with GP practices between 1 Jan 2000 - 30 Jun 2008	1,054,818 women with 735 cancers
Hospital + screening populations									
Goff, 2007, USA	●		●			OC, including borderline	Patient survey	*Cases:* Women with a pelvic mass recruited in secondary care prior to OC diagnosis *Controls:* a) Healthy 'high risk'[b] women enrolled in a screening study [254], b) women who presented for pelvic/abdominal US	*Development-* *Cases:* 74 *Controls:* 243 *Validation-* *Cases:* 75 *Controls:* 245
Andersen, 2008, USA	●			●		OC (NOS)	Patient survey, blood sample	*Cases:* Women with a pelvic mass, recruited prior to OC diagnosis *Controls:* Healthy 'high risk'[b] women enrolled in a screening study[254]	*Cases:* 75 *Controls:* 254
Andersen, 2010, USA	●			●		OC (NOS)	Patient survey, blood sample	*Cases:* Women with a pelvic mass recruited in secondary care prior to OC diagnosis *Controls:* Healthy 'high risk'[b] women enrolled in a screening study,[254] frequency matched to cases on age (</>50years)	*Cases:* 74 *Controls:* 137

Lim, 2012, UK	●		●		●	OC, including borderline	a) Survey, b) Telephone interview, c) GP notes	*Cases:* Women aged 50-79 years with primary OC recruited prior to diagnosis (Feb 2006 – Feb 2008) *Controls:* Screening trial participants,[153] frequency matched on year of birth and agreement to a telephone interview	*Cases:* 194[c] *Controls:* 268[c]
Hospital based population									
Kim, 2009, Korea	●			●	●	Epithelial OC (NOS)	Patient survey, blood sample	*Cases:* OC diagnosis *Controls:* Women with benign ovarian cysts recruited prior to surgery and those undergoing routine Pap smear	*Cases:* 116 *Controls:* 209 (Benign: 74, Pap smear: 135)
Macuks, 2011, Latvia	●			●	●	Epithelial OC (NOS)	Patient survey, blood sample	*Cases:* Women with epithelial OC recruited prior to surgery/diagnosis *Controls:* Age matched 'healthy women' attending a gynaecology outpatient clinic[d]	*Cases:* 24 *Controls:* 31
Shetty, 2015, India	●			●	●	OC, excluding borderline	Patient survey	*Cases:* Women admitted to hospital for investigation and subsequently diagnosed with OC *Controls:* a) Women with benign ovarian pathology, b) those undergoing a 'gynaecological check-up'	*Cases*: 74 *Controls*: 218 (benign: 144, gynaecological check-up: 74)

Jain, 2018, India	●			●	●	OC, excluding borderline	Patient survey, blood sample	*Cases:* Women undergoing surgery for a pelvic mass subsequently diagnosed with ovarian cancer *Controls:* First degree healthy relatives of cases	*Cases:* 45 *Controls:* 90

[a] Data collected on borderline tumours but not included in their tool evaluation.
[b] Women with high-risk family histories consistent with a possible BRCA1 mutation participating in the Ovarian Cancer Early Detection Study (OCEDS).[254]
[c] Numbers varied by study component: questionnaire (191 case, 268 controls), interview (111 cases, 125 controls), GP notes (171 cases, 227 controls).
Abbreviations: OC = Ovarian cancer, Hx = History, NOS = Not otherwise specified, GP = General practice.

5.3.3 Risk of bias

The main potential sources of bias were identified in the "patient selection" and the "index test" QUADAS-2 domains (**Figure 5.2**). Case-control designs can overestimate test performance,[183,184] so we flagged 13 studies as at high ROB for patient selection. In the "index test" domain, key potential sources of bias included failing to pre-define the tool threshold and retrospectively administering the tool after the outcome had been determined e.g. interviewing participants after the ovarian cancer diagnosis had been made. The majority of studies were judged as of low ROB for the "reference standard" and "flow and timing" domains. However, all primary care studies were flagged as at high ROB in the "reference standard" domain as they relied on codes within routinely collected GP data (e.g. the CPRD) to identify ovarian cancer diagnoses, supplemented in two studies by death registration data,[54,242] rather than hospital pathology or cancer registry data. The review question in this study was broad, so concern over the applicability of studies was judged as low. An exception was the "reference standard" domain of a single study which used a composite cancer outcome.[242]

Study	Risk of bias domain			
	Patient selection	Index test	Reference standard	Flow & timing
Lurie, 2009				
Rossing, 2010				
Jordan, 2010				
Hamilton, 2009				
Hippisley-Cox, 2012				
Hippisley-Cox, 2013				
Grewal, 2013				
Collins, 2013				
Goff, 2007				
Andersen, 2008				
Andersen, 2010				
Lim, 2012				
Kim, 2009				
Macuks, 2011				
Shetty, 2015				
Jain, 2018				

Figure 5.2. QUADAS-2 Risk of Bias Assessment.
Green = "Low", Orange = "High", Blue = "Unclear" Risk of Bias.

5.3.4 Tool variables

In total the 16 studies evaluated 21 distinct tools, of which five were diagnostic prediction models.[162] I grouped the variables included in tools into four categories: 1) demographics, 2) personal and family medical history, 3) symptoms, and 4) test results (**Table 5.2**). All tools included symptoms (as per the review inclusion criteria), with 14 including only symptoms. Four tools incorporated demographic variables, two incorporated personal and family medical history, and six incorporated test results.

The most common demographic variable included was age (n=3). This was incorporated as a continuous variable in the QCancer prediction models (n=2) developed by Hippisley-Cox *et al* and as a dichotomised variable (<50 years or ≥50 years) in an ovarian cancer score developed by Grewal *et al*.[54,244] One tool included menopause as a variable.[246] The QCancer models were the only tools to include family history of cancer.

Nineteen different symptoms were incorporated within tools (discounting those in the multi-cancer tool developed by Hippisley-Cox *et al*).[242] Five symptoms (abdominal pain, pelvic pain, distension, bloating, appetite loss) were included in more than half of tools and a further six symptoms (feeling full quickly, difficulty eating, postmenopausal bleeding, urinary frequency, palpable abdominal mass / lump, rectal bleeding) were included in at least a quarter of tools.

Three different types of test result were included within tools: Haemoglobin (n=2), CA125 (n=3) and HE4 (n=2).

Six tools were based on an existing one – the Goff Symptom Index (SI) – which had been modified to include additional demographic, symptom or test result variables. Studies frequently defined variables slightly differently - tool specifications (including the variable definitions used) are summarised in tabular form in **Appendix H**.

Table 5.2. Variables included in final tools.

Tool (Study, year)	Demographics		Personal / Family History		Symptoms													Test results		
	Age	Other	PMH	FH	Abdo. pain	Pelvic pain	Increase Abdo. size / Distens.	Bloat.	Appetite loss	Feeling full	Difficulty eating	Weight loss	Postmen. Bleed.	Rectal bleeding	Palpable abdo. mass / lump	Urinary freq.	Other	Hb	CA125	HE4
Symptom checklists																				
Goff SI (Goff, 2007)					●	●	●	●		●	●									
Modified Goff SI 1 (Kim, 2009)					●	●	●	●		●	●					●	Urinary urgency			
Lurie 7-SI (Lurie, 2009					●	●	●	●	●						●	●	Bowel symptoms, difficulty emptying bladder, dysuria, fatigue, abnormal vaginal bleed.			
Lurie 5-SI (Lurie, 2009)					●	●	●								●	●	Difficulty emptying bladder, dysuria, abnormal vaginal bleed.			
Lurie 4-SI (Lurie, 2009)					●	●	●								●		Abnormal vaginal bleed.			
Lurie 3-SI (Lurie, 2009)							●								●		Abnormal vaginal bleed.			

Hamilton SI (Hamilton, 2009)					●		●	●	●				●	●		●			
SGO consensus criteria[a] (Rossing, 2010)					●	●		●		●						●	Urinary, urgency		
Lim SI 1 (Lim, 2012)					●	●	●	●	●	●		●			●				
Lim SI 2 (Lim, 2012)					●	●	●		●						●		Vaginal discharge		
Hippisley-Cox SI (Hippisley-Cox, 2012)					●		●		●			●	●	●					
Modified Goff SI 2 (Shetty, 2015)					●	●	●	●	●	●	●	●				●	Urinary urgency		
Augmented symptom checklists																			
Goff SI + CA125 (Andersen, 2008)					●	●	●	●		●	●							●	
Goff SI + HE4 (Andersen, 2010)					●	●	●	●		●	●								●
Goff SI + HE4 + CA125 (Andersen, 2010)					●	●	●	●		●	●							●	●

Goff SI + CA125 + menopause (Macuks, 2011)		Meno-pause			●	●	●	●		●	●							●	
Prediction models																			
QCancer Ovarian (Hippisley-Cox, 2012)	●			OC	●		●		●			●	●	●				●	
QCancer Female (Hippisley-Cox, 2013)	●	Townsend score, smoking, alcohol, BMI	T2DM, COPD, endomet. hyperplasia or polyp, chronic pancreatitis	OC, GI cancer, breast cancer	●		●		●			●	●	●			Difficulty swallowing, heartburn / indigestion, blood in urine, blood in vomit, blood when cough, irregular menstrual bleeding, vaginal bleeding after sex, breast lump, breast skin tethering or nipple discharge, breast pain, lump in neck, night sweats, venous thromboembol-ism, CIBH, constipation, cough, unexplained bruising	●		

OC Score A (Grewal, 2013)					●		●	●	●				●	●		●			
OC Score B (Grewal, 2013)					●		●	●	●				●	●		●			
OC Score C (Grewal, 2013)	●				●		●	●	●				●	●		●			

[a] Consensus statement released by the Society of Gynaecologic Oncologists (SGO), the Gynaecologic Cancer Foundation and the American Cancer Society. The terms used to describe a given symptom varied subtly between studies - full details of each tool, including symptom terminology and duration and frequency criteria, are included in **Appendix H.** Abbreviations: PMH = past medical history, FH= family history, Abdo. = abdominal, Distens. = distension, Bloat. = bloating, Postmen. = postmenopausal, Bleed. = bleeding, Freq. = frequency, Hb = haemoglobin, CA125 = cancer antigen 125, HE4 = human epididymis protein 4, SI = symptom index, OC = ovarian cancer, BMI = body mass index, Endomet. = endometrial, T2DM = type 2 diabetes mellites, COPD = chronic obstructive pulmonary disease, GI = gastrointestinal, CIBH = change in bowel habit.

5.3.5 Evaluation of tool performance

Table 5.3 summarises the reported diagnostic accuracy metrics of the tools, the source of accuracy metrics (apparent performance *vs* internal validation *vs* external validation), and the recruitment setting (population *vs* primary care *vs* hospital *vs* hospital + screening) by tool category (symptom checklist *vs* augmented checklist *vs* prediction model).

Most studies calculated the diagnostic performance of a tool directly from the patient sample in which the tool was developed (apparent performance) or by applying internal validation methods such as splitting the sample into development and validation sets (internal validation). Four tools – the Society of Gynaecologic Oncology (SGO) consensus criteria, Goff SI, QCancer Ovarian, Modified Goff SI 1 – were independently validated in an external dataset. The combination of Goff SI and CA125 was evaluated in several studies, but the CA125 thresholds used varied, so no particular combination was externally validated. There was some overlap in evaluation of tools between healthcare settings, but no tool evaluated in primary care was evaluated in another setting or *vice versa*.

The most widely studied tool was the Goff SI, which was evaluated in nine studies.[215,240,241,245,246,248–251] However, two of these studies used subsets of women from the original development study.[249,250] Deviations from the original Goff SI, in how variables were defined, were noted in several studies (**Appendix H**). The only tool to be externally validated in more than one setting was the Goff SI.

5.3.6 Tool diagnostic accuracy

In this section, I compare the accuracy of diagnostic tools. As discussed in **Chapter 1**, the population in which tests and tools are evaluated can affect their diagnostic performance. Therefore, I compared the accuracy of tools by study setting.

Table 5.3. Tool diagnostic accuracy metrics.

Tool (threshold)	Study	Recruitment				Source of accuracy estimate			Sensitivity (95% CI)	Specificity (95% CI)	PPV	AUC (95% CI)
		Population level	1° care	Hospital + screening	Hospital	Apparent performance	Internal validation	External validation				
Symptom checklists												
Goff SI	Goff, 2007			●			●		≥50 yrs: 66.7 < 50 yrs: 86.7	≥50 yrs: 90 <50 yrs: 86.7	-	-
	Andersen, 2008[a]			●			●		64 (52.1-74.8)	88.2 (83.6-91.9)	-	-
	Kim, 2009				●			●	56.9	87.6	-	-
	Rossing, 2010	●						●	67.5 (65.4-69.6)	94.9 (93.9-95.8)	0.77-1.12[b]	-
	Jordan, 2010	●						●	68.1 (65.5-70.7)	85.3	0.09[c] (≥55yrs: 0.21-0.31 <55yrs: 0.04)[d]	-
	Andersen, 2010[a]			●			●		63.5 (51.5-74.4)	88.3 (81.7-93.2)	-	-
	Macuks, 2011				●			●	83.3	48.3	-	-
	Jain, 2018				●			●	77.8	87.8	-	-
	Lim, 2012			●				●	61.4-75.7[e]	89.6-98.9[e]	-	-
Modified Goff SI 1	Kim, 2009				●	●			65.5	84.7	-	-
	Shetty, 2015				●			●	71.6	88.5	-	-
7-symptom Index	Lurie, 2009	●				●			85	40	-	-
5-symptom Index	Lurie, 2009	●				●			80	63	-	-
4-symptom Index	Lurie, 2009	●				●			74	77	-	-
3-symptom Index	Lurie, 2009	●				●			54	93	-	-

Hamilton SI	Hamilton, 2009		●			●			85	85	-	-
SGO consensus criteria	Rossing, 2010	●						●	65.3 (63.1-67.4)	93.9 (92.8-95)	0.63-0.92[b]	-
	Jordan, 2010	●						●	71.5 (69-74.1)	82.9 (81-84.8)	0.08[c] (≥55yrs: 0.18-0.27 <55yrs: 0.05)[d]	-
Lim SI 1	Lim, 2012			●			●		69.6-91[e]	76-91[e]	-	-
Lim SI 2	Lim, 2012			●			●		67.3-91[e]	82.4-94[e]	-	-
Hippisley-Cox SI	Hippisley-Cox, 2012		●				●		71.9	82.9	0.5	-
Modified Goff SI 2	Shetty, 2015				●	●			77	88.5	-	-
Augmented symptom checklists												
Goff SI or CA125[f]	Andersen, 2008			●		●			89.3 (80.1-95.3)	83.5 (78.3-87.8)	-	-
Goff SI or CA125 (>35 U/ml)	Jain, 2018				●	●			97.8	68.9	-	-
Goff SI & CA125 (>21 U/ml)	Macuks, 2011				●	●			79.1	100	-	-
Goff SI & CA125 (>35 U/ml)	Macuks, 2011				●	●			70.8	100	-	-
Goff SI & CA125 (>65 U/ml)	Macuks, 2011				●	●			70.8	100	-	-
Goff SI or CA125[f]	Andersen, 2010			●		●			91.9 (83.2-97)	83.2 (75.9-89)	-	-
Goff SI or HE4[f]	Andersen, 2010			●		●			91.9 (83.2-97)	84.7 (77.5-90.3)	-	-

Any 1 of 3 (Goff SI or CA125 or HE4)[f]	Andersen, 2010			●		●			94.6 (86.7-98.5)	79.6 (71.8-86)	-	-
Any 2 of 3 (Goff SI or CA125 or HE4)[f]	Andersen, 2010			●		●			83.8 (73.4-91.3)	98.5 (94.8-99.8)	-	-
Goff SI & 1 or more of CA125 or HE4[f]	Andersen, 2010			●		●			58.1 (46.1-69.5)	98.5 (94.8-99.8)	-	-
Goff SI & CA125 (>25 U/ml) & menopause	Macuks, 2011				●	●			50	100	-	-
Goff SI & CA125 (>35 U/ml) & menopause	Macuks, 2011				●	●			45.8	100	-	-
Goff SI & CA125 (>65 U/ml) & menopause	Macuks, 2011				●	●			45.8	100	-	-
Prediction models												
QCancer Ovarian (Top 10% risk)	Hippisley-Cox, 2012		●				●		63.2	90.8	0.8	084 (0.83-0.86)
	Collins, 2013		●					●	64.1	90.1	0.5	0.86 (0.84-0.87)

QCancer Ovarian (Top 5% risk)	Hippisley-Cox, 2012		●				●		42.2	95.6	1.1	-
	Collins, 2013		●					●	43.8	95	0.6	-
QCancer Ovarian (Top 1% risk)	Hippisley-Cox, 2012		●				●		13.9	99.3	2.1	-
QCancer Ovarian (Top 0.5% risk)	Hippisley-Cox, 2012		●				●		11	99.6	3.2	-
QCancer Ovarian (Top 0.1% risk)	Hippisley-Cox, 2012		●				●		3.9	99.9	5.5	-
QCancer Female (Top 10% risk)	Hippisley-Cox, 2013		●				●		61.6	90	0.6	0.84 (0.82-0.86)
OC Score A (Score ≥3)	Grewal, 2013		●			●			58.5	97.3	-	0.89
OC Score A (Score ≥4)	Grewal, 2013		●			●			57.6	97.3	-	
OC Score B (Score ≥3)	Grewal, 2013		●			●			75	90.1	-	0.89
OC Score B (Score ≥4)	Grewal, 2013		●			●			58.9	97.3	-	
OC Score C (Score ≥3)	Grewal, 2013		●			●			85.4	85.1	-	0.88
OC Score C (Score ≥4)	Grewal, 2013		●			●			72.6	91.3	-	

[A] Study used a subset of patients from Goff, 2007. [b] Calculated using external data from screening studies.[152,255] [c] Calculated using external Australian population level data. [d] Calculated using external data from US and UK screening studies and Australian population level data.[256,257] [e] Sensitivity and specificity varied by data collection method (Questionnaire, Telephone, GP notes). [f] Biomarker level (CA125, HE4) dichotomised at 95th percentile in control group – levels above that deemed abnormal.

Hospital setting

All but two tools evaluated in hospital populations (including hospital + screening populations) incorporated the Goff SI. Two of these were externally validated – the Goff SI itself and a modified version of the index which incorporated additional symptoms (Modified Goff SI 1). The Goff SI was externally validated in six studies and demonstrated sensitivities which ranged from 56.9% to 83.3% (an outlier result) and specificities from 48.3% (an outlier result) to 98.9%. The Modified Goff SI 1 had a sensitivity of 71.6% and a specificity of 88.5% in a single external validation study.

Augmenting symptom checklists with baseline risk factors and test results generally led to a reduction in sensitivity and increase in specificity, or *vice versa,* depending on the threshold used. For example, the addition of the serum ovarian cancer biomarker CA125 to the Goff SI by Anderson *et al* (2008) led to a reduction in tool sensitivity if both variables were required to be abnormal for a positive tool result. If only one was required to be abnormal for a positive tool result, specificity was reduced.[249]

Population setting

In women recruited from the population setting, two symptom checklists were externally validated side by side – the Goff SI and SGO consensus criteria. The sensitivities and specificities of the tools differed between the studies. However, within each study they were similar, with an in-study maximum difference in sensitivity of 3.4% and in specificity of 2.4% between the tools.

The two population-based case-control studies (Rossing *et al* and Jordan *et al*) used external disease prevalence figures (sourced from screening studies and population level statistics) to estimate PPVs for the Goff SI and SGO consensus criteria, if they were applied in the general population. The tools had similar estimated PPVs within each study, but PPVs were higher in Rossing *et al* (0.63-1.12%) than in Jordan *et al* (<55 years: 0.04-0.05%, ≥55 years: 0.18-0.31%).

Primary care

A single tool (QCancer Ovarian), which took the form of a prediction model and combined symptom variables with demographics, family history and routine blood test results,

underwent external validation in a primary care setting. When the threshold for abnormality was set to include the 5% of women at the highest predicted risk of cancer, QCancer Ovarian had a sensitivity of 43.8% and specificity of 95%. When the threshold was set to include women at the 10% highest risk, the sensitivity increased to 64.1% but specificity fell to 90.1%. Several scores, developed by Grewal *et al*, demonstrated higher sensitivities and specificities than QCancer Ovarian at the 5% risk threshold (OC Score B ≥4) and 10% risk threshold (OC Score C ≥4), but these diagnostic accuracy measures were calculated from the same dataset used in score development.

PPVs were reported for QCancer tools at a range of thresholds (**Table 5.3**). The PPVs at any given risk threshold were similar in the three studies evaluating QCancer tools. For example, values ranged from 0.5-0.8% when the threshold was set to identify the 10% of women at highest risk.

Discrimination was reported for 5 tools - all had similar AUCs within the 'good' range (0.84-0.89). QCancer Ovarian exhibited an AUC of 0.86 on external validation. Tool calibration was assessed for QCancer tools by graphically comparing the predicted cancer risk at two years with the observed risk by predicted risk deciles.[54,242,243] The authors reported that the models showed good calibration on internal validation. On external validation, QCancer Ovarian had reasonable calibration but over-predicted risk, particularly in older women.[243]

5.4 Discussion

5.4.1 Summary

This is the first systematic review to compare the diagnostic performance of symptom-based tools for ovarian cancer detection. I identified 16 studies that had evaluated 21 tools designed to help detect women with undiagnosed ovarian cancer. These tools comprised simple symptom checklists, checklists which included symptoms, demographics and tests, and multivariable diagnostic prediction models. The diagnostic performance of the majority of tools was evaluated solely within the datasets in which they were developed. Four tools were independently externally validated; one in multiple population settings. On external validation, these four tools demonstrated similar (moderate) diagnostic accuracy.

Several checklists, developed in secondary care populations, contained CA125. None of these had been evaluated in primary care. This review did not identify any prediction models incorporating CA125 alongside symptoms.

The variables incorporated within the tools identified in this review were considered as possible candidate variables when I developed a primary care diagnostic prediction model for ovarian cancer (**Chapter 6**).

Key study findings are summarised against the relevant thesis objective in **Box 5.1.**

vii) To perform a systematic review to identify published symptom predicated tools for ovarian cancer detection and to compare these tools in terms of a) included variables and b) diagnostic accuracy

- This systematic review identified 16 studies evaluating 21 symptom-based tools
- Studies were conducted in different settings / populations:
  - The general population
  - Primary care
  - Secondary care +/- screening populations
- Three types of tool were identified:
  - Symptom checklists
  - Checklists of symptoms + risk factors or tests (augmented symptom checklists)
  - Diagnostic prediction models
- Tools included various combinations of:
  - Patient demographics
  - Personal / family medical history
  - Symptoms
  - Tests
- Three symptom checklists and one prediction model were externally validated and demonstrated similar (moderate) accuracy
- No prediction models or primary care tools incorporating CA125 were identified

Box 5.1. Key results against thesis objective vii.

5.4.2 Limitations

I chose to keep the review question and inclusion criteria broad in order to identify all relevant tools, and their constituent variables, in order to inform the development of my prediction model (**Chapter 6**). However, the identified studies were extremely heterogeneous in their design, populations, variable definitions, outcome definitions and thresholds, which ultimately precluded any meaningful meta-analyses. For example, although the Goff SI was evaluated in nine studies there was marked overlap between the participants in three studies; control groups ranged from apparently healthy general population participants to hospital gynaecology patients (with or without benign pathology); ovarian cancer definitions differed; and deviations in the parameters of the SI itself (e.g. required symptom duration and frequency criteria) were noted. While I chose not to perform any meta-analysis, the results demonstrate how the Goff SI performs in different circumstances.

All included studies were at high risk of bias in at least one QUADAS-2 domain, which limits the conclusions that can be drawn.

5.4.3 Tool variables

All of the tools in this review were symptom-based, but they varied markedly in the symptoms they included. This mirrors discrepancies in the literature and within national guidelines, as described in **Chapter 1**, as to which symptoms are associated with ovarian cancer. Some variation is also to be expected given the diverse methodologies and populations used to develop the different tools. Despite this, the symptoms with the highest positive likelihood ratios for ovarian cancer in a recent systematic review (distension, bloating, abdominal or pelvic pain) were incorporated into all but a few tools.[89] Invariably, the more cancer-associated symptoms that are included in symptom checklists, the higher the sensitivity of the tool, but this comes at a cost to specificity as demonstrated by several studies which compared different symptom combinations.[239,245,251] This was cited by Goff *et al* as a rationale for not including urinary symptoms in the Goff SI.[215] However, minor variation in symptoms often had limited impact on performance. For example, on external validation, two studies reported similar diagnostic accuracy metrics for the Goff SI and SGO criteria (which differed on several symptoms). Also, on internal validation, Lim *et al* concluded that changing several of the symptoms made relatively little difference to tool diagnostic accuracy.[251]

In multiple studies, symptom checklists were augmented by ovarian cancer biomarkers with the aim of improving tool diagnostic accuracy. This approach naturally led to a reduction in tool specificity (where *either* symptoms or an abnormal test resulted in a positive tool) or sensitivity (where symptoms *and* an abnormal test were needed for a positive tool). This highlights a major limitation of checklist tools. If ovarian cancer biomarkers are to be included alongside symptoms or other variables within tools, this loss of performance could be avoided by incorporating them within prediction models as per the inclusion of anaemia in QCancer Ovarian. Also, a prediction model approach allows for the inclusion of biomarkers as continuous variables, the importance of which is discussed in **Chapter 3**.

5.4.4 Tool performance

Variation in the reported sensitivity and specificity of the most widely evaluated tool, the Goff SI, was noted between studies. This variation is likely to be due in part to the marked differences in study design, populations and outcome definitions. Despite these differences, in five of the six external validation studies (including two large population-based studies) the Goff SI had a sensitivity in excess of 60% and in all but the smallest study, which included only 24 ovarian cancers and 31 controls, its specificity exceeded 85%. The sensitivities and specificities of the two other externally validated symptom checklists – the SGO consensus criteria and the modified Goff SI 1 - were similar. Give this, it seems unlikely that significant improvements in the diagnostic accuracy of checklists (or models) will be achieved by developing new tools with minor modifications to the symptoms. However, demographic and test variables did add to, and were retained in, QCancer models. It is possible that the performance of these models might be improved by incorporating further predictive demographic or test information.

5.4.5 Clinical relevance

Although a couple of studies included in this review evaluated the diagnostic accuracy of checklists incorporating ovarian cancer biomarkers (CA125 and / or HE4), most of the identified tools were instead developed in order to help select patients for testing. Such tools could be used to identify women at increased risk of undiagnosed ovarian cancer for CA125 testing in primary care. The CA125 predicted models I have developed and present within this thesis could then be used to further inform clinical decision making.

Several of the tools identified in this review are already available for use within the clinical setting in the form of eCDS tools. For example, QCancer tools are integrated within some UK general practice IT systems (EMIS software) and they can alert the clinician if the risk of ovarian cancer in an individual reaches a certain level. eCDS tools have been shown to improve practitioner performance and patient care in some circumstances.[258,259] However, there are multiple barriers to their implementation and they do not always lead to improved outcomes.[259,260] In addition to accuracy (the focus of this review), a tool's cost effectiveness, acceptability to patients and clinicians, and impact on timely ovarian cancer detection and survival need to be evaluated. Currently, a large clustered randomised control trial is seeking to help address this by investigating the clinical impact of implementing a suite of electronic cancer risk assessment tools (including an electronic version of the Hamilton ovarian SI) in UK general practice.[166] Studies have also sought to evaluate the impact of using tools as part of 'proactive symptom triggered testing' programs i.e. where women are actively screened using the tool, with further testing for ovarian cancer occurring if the tool is positive.[‡‡] In one study in the United States, 5,000 women were approached in primary care clinics and screened for symptoms using the Goff SI, with further investigations performed if the Goff SI was positive.[172] However, conclusions were limited as only 2 ovarian cancers were identified in the study window. The DOvE trial also employs a proactive symptom triggered testing approach, supported by media campaigns, in which women can self-refer and are screened for a range of symptoms prior to study inclusion. Although the final DOvE results are yet to be published, a pilot study reported that participants had lower tumour burden and more resectable disease than women diagnosed via the standard clinical pathway.[55]

When considering the clinical utility of a tool it is important to assess the proportion of women who are 'tool positive' who ultimately have ovarian cancer i.e. the PPV. The primary care cohort studies in this review indicated that between 1 in 100 and 1 in 200 women who were QCancer tool positive (5% or 10% risk thresholds) had the disease. Although these figures may appear low, evidence indicates that some patients would opt for cancer testing in clinical presentations with PPVs below 1%.[200] Further, having a positive tool result in the

[‡‡] Such approaches are often referred to as 'symptom triggered screening' in North America (see Goff *et al* 2013[339])

clinical setting does not necessarily mean that further investigation will automatically occur, as there may be a clear alternative cause for the symptoms - the tool is simply intended as a diagnostic aid to highlight the risk of ovarian cancer to the clinician. In proactive symptom triggered testing programs the tool is more than just a diagnostic aid - it is the initial triage step which will dictate whether further ovarian cancer tests take place. The two population studies reporting PPVs relied on external ovarian cancer prevalence figures, but their PPV estimates were similar to that reported in the pilot DOvE study (0.76% in women ≥50 years).[55] Further research is needed to help determine whether, given this PPV, follow-up testing in proactive symptom triggered testing programs is acceptable to women and improves outcomes.

5.5 Conclusions

Over 20 symptom-based tools, incorporating a wide variety of variables, have been developed to help identify women with ovarian cancer. Four tools – the Goff SI, a modified version of the Goff SI, SGO consensus criteria and QCancer Ovarian – have undergone independent external validation and exhibit similar sensitivities and specificities. These tools could play a role in selecting women for CA125 testing, but their acceptability, cost effectiveness and clinical impact need to be evaluated. While checklists including CA125 were identified, these either had poor sensitivity or specificity depending on the threshold used to indicate a 'positive' tool result. No prediction models were identified which included CA125.

Chapter 6. The development and internal validation of ovarian cancer diagnostic prediction models for use in symptomatic women in primary care

6.1 Introduction

During my doctoral research I sought not only to evaluate how well CA125 performed, but to develop novel approaches to augment its performance. Studies have shown that models which incorporate a test result alongside other variables associated with a disease can have greater diagnostic accuracy than the test alone. For example, prediction models which incorporate PSA, demographics, clinical examination findings and additional test results have been shown to significantly outperform PSA alone in the detection of prostate cancer in the screening setting.[169] **In Chapter 3**, I described the development of a model combining CA125 level and patient age to estimate the probability of ovarian cancer. As this model contained only two variables, it had the advantage of simplicity. However, the predictive ability of models tends to improve when multiple important predictive variables are included. Ewout Steyerberg, one of the world's leading experts on prediction models, recommends considering between 5 and 20 variables when developing a clinical prediction model.[171] In the systematic review presented in **Chapter 5**, I found that a range of variables including patient demographics, family and personal medical history, symptoms and test results, had been incorporated into tools used for ovarian cancer detection. I postulated that a model incorporating these types of variables alongside CA125 would outperform CA125 alone in the detection of ovarian cancer in symptomatic women undergoing testing in primary care.

When developing a new prediction model, it is important to evaluate its overall performance in terms of its discrimination and calibration. It is also important to consider its diagnostic accuracy at clinically relevant thresholds, as thresholds are invariably used to guide management when models are implemented in clinical care pathways. For example, NICE recommends offering high statin treatment if the QRisk model indicates that a patient's 10-year risk of developing cardiovascular disease is ≥10%.[164]

In this chapter, I describe the development of two multivariable diagnostic prediction models intended for use in symptomatic women undergoing CA125 testing in primary care. I detail their overall performance on internal validation. Given the importance of selecting the most appropriate 'action thresholds' for use in clinical practice, I also describe the accuracy of different model-derived risk-based thresholds. Finally, I explore the potential diagnostic implications of applying these thresholds in primary care, using my study cohort as an exemplar.

This study addresses my final two thesis objectives:

viii) *To develop and internally validate an ovarian cancer diagnostic prediction model incorporating symptoms, test results (including CA125), and risk factors.*

ix) *To explore the potential diagnostic implications of implementing 'action thresholds', based on prediction model derived estimated ovarian cancer probabilities, within primary care.*

The following paper, based on the work described within this chapter, has been submitted for publication and, at the time of thesis submission, is under review:

Could ovarian cancer prediction models improve the triage of symptomatic women in primary care? A modelling study using routinely collected data. Garth Funston, Gary Abel, Emma J. Crosbie, Willie Hamilton, Fiona M. Walter.

6.2 Methods

6.2.1 Sample size considerations

This study was performed using data obtained to undertake the CA125 diagnostic accuracy study described in **Chapter 3**. In order to maximise the sample size, I used data from all available patients who met selection criteria (**Section 6.2.5**). No formal sample size calculation was performed prior to analysis.

New recommendations for calculating minimum sample sizes for clinical prediction modelling studies have recently been published (after this study was commenced) by Riley *et al*.[261] They

suggest an approach to calculating the minimum sample size required for modelling studies based on the number of predictors being considered, the number of participants, the disease incidence and the expected model performance. However, traditionally, the number of Events Per Variable (EPV) has been considered the key factor in ensuring adequate sample size for prediction model studies.[162,171] The EPV is the number of outcomes (in the case of this study, ovarian cancers) per predictor being considered for model inclusion (known as 'candidate predictors').[171] More strictly, EPV calculations should take account of the total number of degrees of freedom from candidate variables.[262] For example, if ethnicity is categorised into five groups, it would have 4 degrees of freedom and so should be thought of as 4 'variables' in EPV calculations. Any other terms included during the model development processes, such as interaction and polynomial terms, should also be considered in EPV calculations. *Too low* an EPV risks 'overfitting' the model to the derivation sample, so that it performs well within that sample but poorly when evaluated in another sample.[171] Importantly, Steyerberg recommends that any variable which is chosen based on patterns within the data should be counted within the EPV calculations.

There is significant debate in the literature with regard to the recommended minimum number of EPVs for model development studies. The most widely quoted figure is 10 EPVs, but some simulation studies indicate that this may be too stringent, with 5 EPVs suggested as a minimum.[263,264] While the study sample size and number of outcomes was dictated by available data, I was mindful of the EPV guidance when identifying variables for model development and aimed for an EPV close to 10, with an absolute minimum of 5.

6.2.2 Model variables

In his highly influential book on clinical prediction models, Steyerberg comments that model specification is "*the most difficult part of prediction modelling*" and that *"it is virtually impossible to obtain a reliable answer to the question: which predictors are important and which are not?".*[171] This neatly sums up one of the key challenges I faced in this study: which variables, from the myriad recorded within the CPRD, should I include as variables in my models? Steyerberg recommends that clinical expertise and evidence from the literature should be used as a first step in selecting variables.[171] After a list of candidate variables are selected based on available evidence and experience, there are two options: include all of the

identified variables in the model (full model approach), or use data driven methods to select the final variables for model inclusion from the list of candidates. Data driven approaches are useful in removing variables which have little or no predictive effect within the model. Reducing the number of variables within a model may also improve clinical utility as less information needs to be collected when using them.[171] I developed two models within this study: Model 1 and Model 2. Model 1 was pre-specified (full model approach) and consisted of age and CA125. This model is analogous to the age and CA125 model presented in **Chapter 3**. Model 2 contained age, CA125 and other variables associated with ovarian cancer risk. To determine which variables to include in Model 2 I used a two-step approach:

1) Selection of candidate variables based on expertise and the literature (pre-modelling)

2) Selection of final model variables using data driven selection procedures (during modelling)

In the next two sections I describe the first step in variable selection for Model 2. Selection of the final variables for Model 2 using data driven procedures is described within **Section 6.2.8**.

6.2.3 Identifying possible candidate variables

The first step I took in identifying candidate variables was to carry out a systematic review of ovarian cancer detection tools, described in **Chapter 5**. This identified three broad categories of variable: 1) symptoms, 2) tests, and 3) baseline risk factors (demographics, personal and family medical history). I considered variables in any of these categories for Model 2. However, few of the included studies were explicit about how candidate variables were selected and, for those which were explicit, few appeared to have followed a systematic approach in their selection. It was not feasible to perform a systematic review, or series of reviews, covering all possible risk factors, tests and symptoms of ovarian cancer in order to inform variable selection in this study. Instead, I chose to supplement my systematic review of ovarian cancer tools with information from high level evidence sources, to help ensure that key predictors were not missed from my model (**Box 6.1**). Using these sources, I developed a provisional list comprising more than 40 candidate variables. This list was reviewed by the multidisciplinary research team consisting of experienced GPs (Prof Willie Hamilton, Prof

Fiona Walter), a statistician (Prof Gary Abel) and a gynaecological oncologist (Prof Emma Crosbie). The final list of candidate variables was agreed upon at a consensus meeting. The following questions were used to help select candidate variables from the provisional list:

1) Is the variable routinely recorded within primary care records / is there likely to be significant missing data?
2) Is there reasonable evidence for an association with ovarian cancer e.g. from systematic reviews and meta-analyses?
3) Is the variable very similar to other variables that warrants inclusion? If so, would it be appropriate to include only one?

Symptoms

NICE guidelines (CG122 and NG12)[57,65]

Tests

Key primary care research studies and systematic reviews

Baseline risk factors

World Cancer Research Fund / American Institute for Cancer Research reports on "Diet, nutrition, physical activity and ovarian cancer"[274]

International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans[41]

Cancer Research UK: Ovarian cancer risk factor statistics[278]

Key narrative review articles[25,319,322,340]

Key systematic reviews and meta-analyses

Box 6.1. Additional information sources consulted to identify candidate variables.

Excluded variables

A list of variables which were considered, but ultimately rejected as candidates, is included in **Appendix I** alongside the rational for their exclusion. Two excluded variables warrant further

discussion due to their strong associations with ovarian cancer: germline BRCA mutations and a family history of BRCA-associated cancers.

As discussed in **Chapter 1**, germline BRCA mutations significantly increase the risk of ovarian cancer in carriers: the cumulative risk for BRCA 1 carriers by age 70 has been estimated at 39-59%.[20,21] Risk of ovarian cancer is also greater in women with family history of breast or ovarian cancers, with risk increasing if multiple first degree relatives are affected at a young age or at multiple sites.[25,26] Of note, family history of ovarian cancer was included in the QCancer Ovarian prediction model (**Chapter 5**).

Despite their importance as risk factors for ovarian cancer, I harboured concerns over the quality of the recording of these variables within primary care records. I developed Read code lists (**Appendix C**) for each variable using the code browser supplied by the CPRD. I used these lists to identify women in my baseline cohort who had a relevant code within CPRD (recorded prior to CA125 testing) and who were subsequently diagnosed with ovarian cancer. I compared the proportions of women with relevant codes to figures drawn from the literature (**Table 6.1**). No BRCA mutations were recorded in women prior to CA125 testing and the proportions of women with a coded family history of ovarian and breast cancer was much lower than published figures. While this comparison has limitations (due to differences in the populations, study design and outcome definitions) it did support the assertion that there is likely significant missing data pertaining to BRCA status and family history of cancer in the CPRD. In addition to poor coding of family cancer history, I felt it was also possible that it would be subject to 'recording bias', with GPs more likely to ask about and code family history if they had a strong suspicion of ovarian cancer.

Table 6.1. Comparison of the proportion of ovarian cancer cases in the baseline cohort who had codes for a) germline BRCA mutation and b) family history of breast and ovarian cancer, with figures drawn from the literature (case-control studies).

Variable	CA125 cohort, %	Published studies, %
Germline BRCA mutation	0	10-15[25]
Family history of ovarian cancer	1.2	2.6-6.2[265–267]
Family history of breast cancer	3.4	7.3-11.9[265–267]

6.2.4 Included candidate variables

19 candidate variables were chosen for inclusion in data driven variable selection procedures (**Box 6.2**). In this section, I summarise the rationale for their inclusion.

Symptoms

As discussed in **Chapter 1**, studies have shown that certain symptoms are more common in women with ovarian cancer than in healthy controls. Symptoms have previously been included in primary care ovarian cancer prediction models e.g. QCancer Ovarian (**Chapter 5**). However, not all symptoms of ovarian cancer are equally predictive. For example, the most recent systematic review of ovarian cancer symptoms indicates that abdominal distension is more indicative of ovarian cancer (higher positive likelihood ratio) than other symptoms such as abdominal pain, appetite loss and fatigue.[89] In addition, studies have shown that the PPVs for combinations of symptoms are higher than for individual symptoms. Given this, I posited that the type of symptom(s) a woman has might contribute to the diagnostic performance of a CA125 based prediction model.

Symptoms

- Abdominal / pelvic pain
- Appetite loss
- Bloating
- Distension
- CIBH
- Fatigue
- Urinary frequency / urgency
- Weight loss
- New IBS (≥50 years)

Blood biomarkers

- CA125
- Albumin
- Haemoglobin
- Platelets
- CRP

Risk / protective factors

- Age
- Ethnicity
- Height
- BMI
- Personal history of breast cancer

Box 6.2. Model 2 candidate variables.
CIBH = change in bowel habit, IBS = irritable bowel syndrome, CRP = C-reactive protein, BMI = body mass index

In the review presented in **Chapter 5**, 19 different symptoms were included within the various ovarian cancer tools. Given the limitations on number of candidate variables, I could not include all of these symptoms. Ultimately, I chose the symptoms listed in current NICE guidelines as candidates. Most patients who underwent CA125 testing in my cohort are likely

to have done so due to one of these symptoms. The NICE list is broad and includes the symptoms most strongly associated with ovarian cancer in a recent systematic review: abdominal distension, abdominal and pelvic pain, bloating and loss of appetite.[89] Pragmatically, using the NICE symptom list might allow any models developed in this study to be more easily incorporated into the current diagnostic pathway.

New diagnoses of Irritable Bowel Syndrome (IBS) in women ≥50 years is unusual and may represent misattribution of symptoms such as bloating and change in bowel habit which are due to ovarian cancer.[92,268] Whilst new IBS does not represent a single symptom, it is discussed alongside symptoms in NICE guidelines and I have included it in the symptom category in this study for consistency.

Blood tests

In addition to CA125, two other blood tests were included in the ovarian cancer tools identified in my systematic review: HE4 and haemoglobin. HE4 is a promising ovarian cancer biomarker (I am currently conducting a prospective pilot study comparing its performance against that of CA125 in primary care). However, it is not available within English primary care and so could not be included as a candidate variable. By contrast, haemoglobin level (as part of a full blood count) is one of the most commonly requested tests in primary care.[269] A low haemoglobin level was a predictor of ovarian cancer (2.3 fold increased risk) in the QCancer ovarian model and so I included haemoglobin level as candidate variable.

Several other routinely performed blood tests have emerged as primary care cancer biomarkers in the last few years, most notably platelet count, albumin level and inflammatory markers. A large primary care case control study reported that the 2-year incidence of cancer was significantly greater in primary care patients with a raised platelet count than a normal platelet count.[174] Ovarian cancer was the 4th most common cancer in the female thrombocytosis group within that study. Low albumin is common in women with ovarian cancer in the secondary care setting.[270] In addition, a retrospective primary care cohort study has reported an association with hypo-albuminaemia and cancer diagnosis (within 12 months of testing).[271] In 2019 a large primary care case control study found that the one-year incidence of cancer was higher in patients with elevations in the inflammatory markers C-

reactive protein (CRP), Erythrocyte Sedimentation Rate (ESR) and Plasma Viscosity (PV) than in patients with normal inflammatory marker levels.[175] While research has predominantly focussed on the value of platelets, albumin and inflammatory markers as biomarkers for cancer in general (rather than specifically for ovarian cancer) I considered that there was sufficient evidence to include them as candidate variables.

CRP, ESR and PV are associated with similar inflammatory processes and there is little difference between the inflammatory markers in terms of their association with cancer in primary care.[175] Therefore, in the interests of limiting the number of candidate variables, I took only CRP (the most common of the three tests within the baseline cohort) forward into data driven variable selection procedures.

Risk / protective factors

- ***Age***

As discussed in earlier chapters, the incidence of ovarian cancer varies markedly by age. As such, age was included as a candidate predictor variable.

- ***Ethnicity***

Ovarian cancer incidence rates in England are lower in women of Asian and Black ethnicity than of White ethnicity.[7,272] Age standardised rates in England (2002-2006) ranged from 17.4-18.1/100,000 for white women, 9.2-15.5/100,000 in Asian women and 6.6-12.1/100,000 in black women. Information on ethnicity was available from both the CPRD and HES APC and, on the basis of previous research, I anticipated only around 3% of women would have a missing ethnicity in both data sources.[273]

- ***Adult attained height***

The 2018 World Cancer research Fund (WCRF) report on diet, nutrition, physical activity and ovarian cancer (which is informed by a series of high quality systematic literature reviews) found convincing evidence that adult attained height is associated with increased risk of ovarian cancer.[274] A number of other systematic reviews and meta-analyses have also found a significant association between increasing height and increased ovarian cancer risk.[36–38]

Height itself is unlikely to be a causative factor in ovarian cancer development but rather a marker for environmental and genetic factors associated with the disease.

- **Body *Mass Index (BMI)***

In their 2014 report on diet, exercise, physical activity and ovarian cancer, the WCRF reported that there was strong evidence that higher body fatness (as measured by BMI) is associated with increased risk of ovarian cancer.[274] This finding has been corroborated by a number of large studies and meta-analyses.[34–37]

- ***Personal history of breast cancer***

As discussed earlier in this chapter, genetic mutations, including germline BRCA mutations, substantially increase the risk of both ovarian and breast cancer. There is also evidence that women with a personal history of breast cancer are at increased risk of subsequent ovarian cancer.[275] While I discounted BRCA mutations and family cancer history as potential variables due to poor data recording, I judged that personal history of breast cancer could be included as historical NCRAS data on all cancers diagnosed within the cohort was available (obtained specifically for this purpose).

6.2.5 Defining the study cohort

I saw symptoms as a key category of candidate variable for the development of Model 2. However, the fact that symptoms are not always coded within CPRD posed a methodological issue - if the baseline cohort were to be used for model development, patients with symptoms recorded within free text who lacked a symptom Read code would be misclassified. This could lead to biased model estimates, particularly if symptoms are more likely to be coded in severe presentations / disease. To mitigate this issue, I applied further restrictions to the baseline cohort, excluding women who lacked a Read code for a symptom of ovarian cancer (as listed in current NICE guidelines) in their CPRD record in the 12 months prior to their index test date.

NICE recommends that women with a pelvic mass or ascites on examination be referred on the urgent cancer pathway rather than be tested for CA125 in primary care, due to the high predictive values of these signs.[89,114] The models developed in this study were intended to be used to help make decisions on further investigation and referral in CA125 tested women. I

did not feel that it would be useful, or appropriate, to use these models in primary care in women who had such strongly predictive features of disease and who already warranted further investigation. I therefore chose to exclude women with a code for a pelvic mass or ascites in their CPRD record in the 12 months prior to the index test date.

I utilised a pre-existing symptom Read code list, kindly provided by Professor Willie Hamilton, to identify symptoms and signs of ovarian cancer in CPRD in order to define my cohort. To ensure that the code list was complete and up-to date, I carried out searches for key symptom terms using the Read code browser provided by the CPRD. I identified a few additional relevant terms (**Appendix C**) which were added to the list. I also identified codes for a diagnosis of IBS (**Appendix C**), to be used in women ≥50 years of age, and included them in the list.

When preparing my list of candidate variables I considered including a history of oophorectomy as a protective factor. However, this was discussed at the consensus meeting and we agreed that including women with bilateral oophorectomies would not be appropriate as the incidence of ovarian cancer in these women is extremely low.[276] For this reason, most of the ovarian cancer detection tool studies identified in the systematic review (**Chapter 5**), specifically excluded women with bilaterally oophorectomies. I followed their example. I developed a Read code list for bilateral oophorectomy, which was reviewed by Professor Emma Crosbie before use (**Appendix C**). This list was used to exclude women with a code for bilateral oophorectomy in their CPRD record at any point prior to the index test date.

6.2.6 Preparation of candidate variables

In this section, I describe the preparation of candidate variables in my study cohort prior to statistical analyses. The data source for each variable, in addition to the period of variable inclusion (e.g. ever *vs* the 12 months pre-index test date) and details of any data categorisation, are summarised in **Table 6.2**. As the models were intended to be used at the point when a CA125 test has been performed and result made available to the GP, I only considered predictor variable data recorded on, or prior to, the index test date. The exception to this was ethnicity, as a patient's ethnicity should remain unchanged. To maximise

completeness and minimise the amount of data which had to be imputed I included ethnicity recorded at any point in a patient's record.

Table 6.2. Preparation of candidate variables.

Variable	Data source	Categorisation	Variable inclusion time/period
Symptoms			
Ovarian cancer symptoms[114]	CPRD	Binary for each symptom. Presence/absence of: abdominal/pelvic pain, appetite loss, bloating, distension, change in bowel habit, fatigue, urinary frequency/urgency, new IBS (≥50 years old), weight loss	In the 12 months prior to CA125 testing
Blood biomarkers			
CA125	CPRD	Continuous	First valid CA125 level in study period
Albumin[271]	CPRD	Categorical: *Not tested* *<35 g/L* *≥35 g/L*	Most recent record on or in the 12 months prior to the CA125 test date
Haemoglobin[54]	CPRD	Categorical: *Not tested* *<12g/dl* *≥12g/d*	Most recent record on or in the 12 months prior to the CA125 test date
Platelets[174]	CPRD	Categorical: *Not tested* *<300x10^9/L* *300-449x10^9/L* *≥450x10^9/L*	Most recent record on or in the 12 months prior to the CA125 test date
CRP[175,277]	CPRD	Categorical: *Not tested* *<3 mg/L* *3-9.99 mg/L* *≥10 mg/L*	Most recent record on or in the 12 months prior to the CA125 test date
Risk / protective factors			
Age[274,278]	CPRD	Continuous (years)	On date of CA125 testing
Ethnicity[7,25]	CPRD and HES	Categorical: *White* *Other ethnicities*	Most frequently recorded[273]
Height[36–38,274]	CPRD	Continuous (cm)	Most recent on/prior to CA125 test date recorded when ≥18 years old
BMI[34,35,274]	CPRD	Continuous (Kg/m^2)	Most recent on/in the 10 years prior to CA125 test date when ≥18 years old
Personal history breast cancer[275]	CPRD / NCRAS	Binary	Up to CA125 test date

Symptoms

The updated ovarian cancer symptom code list was used to identify all relevant coded symptoms in the 12 months prior to the index test date. Each of the nine symptoms were coded as a separate binary variable (present / absent).

Blood tests

While all women in the cohort had a valid CA125 level recorded, not all women had recent tests for platelets, albumin or CRP. I considered imputing missing biomarker levels and including them in my analyses as continuous variables. However, this would have limited the clinical utility of the model as it could then only be used in practice if levels of all biomarkers were known for a patient (unless a series of models were developed). I also considered dichotomising the biomarkers into a) normal or not performed vs b) raised, as Hippisley-Cox *et al* did for haemoglobin level in the QCancer Ovarian model. There were two arguments against this approach. Firstly, dichotomising continuous data leads to a significant loss of information; as was demonstrated in **Chapter 3,** risk can vary substantially with specific test levels above a given threshold. Secondly, in their work on inflammatory markers, Watson *et al* found that patients who had a blood test performed had a significantly higher risk of cancer than those who did not, even if that test was normal.[175] So, I judged that whether or not a woman has had a particular test might be predictive in its own right.

Ultimately, I chose to include all blood tests, save for CA125, as categorical variables. Using patterns within the data to decide upon categories can lead to overfitting and so is not recommended by Steyerberg.[171] Therefore, I avoided data driven categorisation and used information from published studies, existing models and standard reference ranges to categorise each blood test into 3-4 categories (including a no test group). The distributions of test levels were only reviewed to ensure that no group was particularly small (<~2% of the cohort), as small groups can affect the reliability of prediction model estimates.[171]

- ***CA125***

The preparation of CA125 level data has been described previously (**Chapter 2**). The index CA125 level was kept as a continuous variable for analyses.

- ***Platelet count***

Codes for platelet count (**Appendix C**), recorded in the test file of the CPRD on or in the 12 months preceding the index test date, were identified. Where multiple platelet counts existed, the most recent was used. Where multiple platelet counts occurred on the same day, I used the mean.

The standard upper reference range for platelets is 450 x 10^9/L. However, there is evidence that patients with 'high normal' platelet counts in primary care have a greater incidence of cancer than those with 'low normal' counts.[279] This was taken into account when categorising platelets. 4 categories were used:

1) Not tested
2) <300x10^9/L
3) 300-449x10^9/L
4) ≥450x10^9/L

- ***Haemoglobin level***

I identified codes for haemoglobin level (**Appendix C**), recorded in the test file of the CPRD on or in the 12 months preceding the index test date. Entries recorded in g/L were converted to g/dl. Where multiple haemoglobin levels existed, the most recent was used. Where multiple haemoglobin levels were recorded on the same day, I used the mean.

I categorised women into three groups:

1) Not tested
2) <12g/dl
3) ≥12g/dl

- ***Albumin level***

I identified codes for albumin level (**Appendix C**), recorded in the test file of the CPRD on or in the 12 months preceding the index test date. Where multiple levels existed, I used the most recent. Where multiple albumin levels were recorded on the same day, I used the mean. I Initially categorised albumin level in line with Merriel *et al*,[271] with hypalbuminaemia broken down into 'mild', 'moderate' and 'severe'. These groups were small (severe

hypoalbuminaemia [<30g/L] contained only 140 women), so I generated a single hypoalbuminaemia group. The final categories were:

1) Not tested
2) <35 g/L
3) ≥35 g/L

- ***CRP level***

Codes for CRP level (**Appendix C**), recorded in the test file of the CPRD on or in the 12 months preceding the index test date, were identified. Where multiple levels existed, I used the most recent. Where multiple CRP levels were recorded on the same day, I used the mean. Reference ranges for CRP vary depending on the clinical application e.g. assessment for acute infection or as a risk marker for cardiovascular disease. Greater than 3 mg/L is often considered abnormal.[280] I categorised women into the following groups:

1) Not tested
2) <3 mg/L
3) 3-9.99 mg/L
4) ≥10 mg/L

Risk/protective factors

- ***Age***

The preparation of patient age is described in **Chapter 2**. I used age in years at index test date, as a continuous variable, in this study.

- ***Ethnicity***

Ethnicity is recorded within the CPRD using hierarchical codes which map to 1991 and 2001 census categories. I used a Read code list, developed by Mathur *et al*,[273] to identify codes for ethnicity recorded at any point within a woman's CPRD record. These codes can be classified into 16 census categories which further collapse into 5 overarching groups (**Table 6.3**). A patient may have ethnicity coded more than once within their CPRD record and have different

coded ethnicities. I followed the approach taken by Mathur *et al* to determine patient ethnicity from CPRD:[273]

1) If a single ethnic group was recorded, I accepted that group
2) If more than one ethnic group was recorded, I accepted the most common ethnic group
3) If more than one ethnic group was recorded with the same frequency (i.e. a 'tie'), I accepted the most recent ethnic group recorded
4) If there was a tie on the most recent date, I recorded CPRD ethnicity as missing

Not all patients have an ethnicity recorded within the CPRD, but HES APC data provided an additional source of ethnicity information. In HES APC, ethnicity is classified as one of 11 categories which map to the 5 ethnic groups listed in **Table 6.3**. Only one ethnicity is recorded for each patient in HES APC (HES use a similar approach to that used by Mathur *et al* to prepare ethnicity before it is provided to researchers).[192] Where ethnicity data was missing in CPRD, I accepted the HES ethnicity.

Table 6.3. Five group and 16 group ethnic categories.

Five ethnic groups	16 ethnic categories
White	British or mixed British
	Irish
	Other White
Mixed	White + Black Caribbean
	White + Black African
	White + Asian
	Other mixed
Asian	Indian or British Indian
	Pakistani or British Pakistani
	Bangladeshi or British Bangladeshi
	Other Asian
Black	Caribbean
	African
	Other Black
Other ethnic groups	Chinese
	Other

I initially classified ethnicity into the five groups, but the numbers of women in individual ethnic groups, other than White, were small e.g. 160 women (0.56%) were of Mixed ethnicity. So, I further collapsed five group ethnicities into two groups: "White" and "other ethnicity".

Where no ethnicity could be identified, either from CPRD or HES APC data, multiple imputation was used to replace it (**Section 6.2.7**).

- ***Height***

Patient height was identified from the Additional file in the CPRD. As the majority of women reach their maximum height by age 18 years, I used data recorded when women were ≥18years.[281] Heights recorded in metres were converted to cm. Implausible values for height were excluded (<1.21m and >2.14m). The plausible height range used encompasses both the lower and upper 99.6 centiles for women at age 18 years and has previously been used in CPRD research.[282,283] After excluding implausible heights, I identified the most recent height recorded (on or prior to the index test date) for each patient. Multiple imputation was used to replace missing height data (**Section 6.2.7**).

- ***Body Mass Index (BMI)***

I calculated BMI for patients using the most recent plausible weight (excluding <20kg) recorded within the Additional file of the CPRD in the ten years prior to the index test date, and the most recent height (excluding <1.21m and >2.14m) recorded on or prior to the index test date. Plausible ranges were informed by the literature.[283,284] I took the pragmatic approach of excluding weights recorded more than ten years prior to the index test date due to the possibility of significant weight change during that period. Measurements recorded at ages <18 years were also excluded. Where more than one weight or height were recorded on the same day, I used the mean in my BMI calculation. As in other CPRD studies, where no BMI could be calculated directly from weight and height due to missing data, I accepted the most recent directly entered BMI value recorded in CPRD on, or in the ten years prior to, the index test date.[35,283] I excluded implausible BMI measurements (<5kg/m^2 and >200kg/m^2). The choice of a plausible BMI range was informed by the literature.[35,283] Where no BMI could be ascertained, it was replaced by multiple imputation (**Section 6.2.7**).

- ***Personal history of breast cancer***

I identified women with *in situ* or invasive primary breast cancer recorded in either the cancer registry (ICD10 codes C50 and D05) or the CPRD (code list provided by Professor Willie Hamilton, **Appendix C**) on or prior to the index test date. Although there are issues with

recoding of cancer diagnoses in the CPRD (as discussed in **Chapter 2**), I chose to include it as a data source to improve completeness as cancer registry data is only available from 1990 onwards.

6.2.7 Missing data

A proportion of women did not have a recorded ethnicity, height or BMI. I could have excluded patients without complete data for all candidate predictors and have performed a complete case analysis (CCA). However, this would have reduced the sample size and could have introduced a selection bias. An alternative approach to CCA is to impute the missing information into the data set. Simple approaches can be used, such as replacing missing continuous data with the mean value or missing categorical data with the modal group. However, these approaches do not account for correlations between different variables and underestimate variability in the data. Using correlations between variables within a dataset, regression imputation can be used to predict and impute missing data to create a single complete dataset, but this approach fails to capture the inherent uncertainty in the imputed data points. For these reasons multiple imputation (MI) is often the favoured approach to imputation.[171] In MI, an imputation model uses correlations between variables within the dataset to replace missing data multiple times, thereby generating multiple complete datasets. This allows the uncertainty in the imputed values to be captured. Following the imputation step, analysis is performed separately on each imputed dataset and then the results are combined across the datasets. A downside of multiple imputation is that working with multiple datasets complicates data handling and analysis.[285]

When considering which approach to use for missing data (complete case analysis vs MI) it is helpful to explore the potential 'mechanism' by which the data is missing. There are three broad mechanisms of missingness:

1) Missing completely at random (MCAR) – Neither unobserved data nor the missing value predicts missingness

2) Missing at random (MAR) – Other known variables (observed data) predict missingness

3) Missing not at random (MNAR) – The value of the missing variable itself predicts whether it is missing

I explored potential mechanisms of missingness within my data, using logistic regression, and found that missing ethnicity, height and BMI were associated with other candidate variables. This indicates that their mechanism of missingness is not MCAR and so a CCA approach would not have been appropriate, as the retained patients may not be representative of the sample.[171] It is not generally recommended that MI be performed if a NMAR mechanism is suspected (due to the risk of biased imputations) but there is no way to distinguish MAR from MNAR without obtaining the missing data. For the purposes of this study I assumed that missing data were MAR and used a MI approach.

I used multivariate imputation by chained equation (MICE) to impute missing data.[286] MICE allows different models to be chosen for the imputation of different types of variable e.g. linear regression for continuous variables and logistic regression for binary variables. Linear regression was used to impute height and BMI data while logistic regression was used for two-category ethnicity. An imputation model should contain at least as many variables as the prediction model (the principal of congeniality).[171] So, all candidate variables (main effect and non-main effect) and the outcome variable, were included in my imputation model. Where variables were transformed, this was done prior to their inclusion in the MI model. When performing MI, the number of imputations must be specified. Stata recommends performing a minimum of 20 imputations.[287] Some authors recommend that the number of imputations should be similar to the proportion of patients with missing data.[288,289] 14.2% of women in my study cohort had missing data for one or more variables, so I chose to perform 20 imputations thereby generating 20 'complete' datasets. Following imputation, I examined the means, standard errors and ranges of imputed BMI and height data and the categories of ethnicity, confirming that imputed data were reasonable.

6.2.8 Model derivation

Prior to model derivation, I mean-centred all continuous variables. BMI and CA125 level were right skewed, so were log transformed. As discussed in **Chapter 3**, the relationship between log CA125 and age with ovarian cancer is non-linear, so restricted cubic splines (5-knots) were

once again generated for log CA125 and age.[208] To account for a possible interaction between CA125 level and age I generated an interaction term.

Model 1 (age and CA125 level) was pre-specified and was derived by performing logistic regression on the study dataset. In order to develop Model 2, a logistic regression model containing all predictor variables was fitted to the data using the *mi estimate* command, which applied Rubin's rules to adjust coefficients and standard errors for the variability between the 20 imputed datasets.[287] To select the final variables for inclusion in Model 2, I used a backward elimination approach removing the least significant variable, then refitting the model and repeating until all retained variables had a p value of ≤0.05. The Wald test was used to assess the significance of each categorical variable. There is no consensus within the literature as to which significance threshold to use in data driven variable selection procedures.[171] For small data sets, large p-values are often recommended while for very large samples small p-values are recommended.[171] Given a reasonably large study sample size, I chose to apply the standard p-value. Variable coefficients were used as weights for both models.

For Model 2, I examined 19 candidate variables with a total of 32 degrees of freedom (main effect and non-main effect), giving an EPV of 9.

6.2.9 Discrimination and calibration

In order to assess model discrimination, I calculated the AUC of each model. For Model 2, the AUC was calculated for each imputed dataset and Rubin's rules were used to combine results across imputed datasets. Calculating the AUC directly from the data from which the models were developed provided the apparent performance. However, apparent performance can give an overly optimistic impression of model performance due to overfitting. I used 10-fold cross-validation as a method of internal validation in order to assess for optimism in AUC estimates. In 10-fold cross-validation, the sample is split into ten equal groups, the model is developed in one of these groups and tested on the remaining sample.[171] This is repeated ten times so that each group is used both in development and testing of the model. The mean AUC across the ten iterations is then calculated. I used the *cvAUROC* user written command in Stata to perform this analysis.[290] For Model 2, I calculated the cross-validation AUC for each

imputed data-set then combined these across imputed datasets in accordance with Rubin's rules. The directly computed AUCs were compared with the cross-validation AUCs in order to quantify model optimism.

For the purpose of comparison, the AUC of CA125 was calculated directly from the study cohort using the *Roctab* command in Stata.[206]

As recommended by Van Calster *et al,*[291] I calculated the calibration slope (with 95% confidence intervals) of models on internal validation (10-fold cross validation). I used a version of *cvAUROC* command which had been modified for this purpose by Prof Gary Abel. For Model 2, this code used Rubin's rules to combine the results across imputed datasets in order to give a single slope metric. The calibration slope provides an indication of whether, on average, a model over- or under-estimates risk. The perfect slope is 1. Slope values less than 1 indicate that the model estimated probabilities are too extreme (too high in those at high risk and too low in those at low risk).[291] A slope greater than 1 indicates the reverse.

6.2.10 Thresholds for further investigation

I devised thresholds to identify women with a ≥1%, ≥2% and ≥3% probability of undiagnosed ovarian cancer based on the models. The lowest threshold – ≥1% – was selected as patients have reported that they would opt for cancer investigations at this risk level.[200] The highest threshold – ≥3% – was chosen to match the risk threshold used by NICE in 2015 when making recommendations on urgent specialist cancer investigation or referral for symptomatic primary care patients.[57] I compared the diagnostic accuracy (sensitivity, specificity, PPV and NPV) of these model thresholds with that of CA125 at the conventional cut-off (≥35 U/ml).[114] I compared specificity, PPV and NPV of the model thresholds to that of CA125 cut-offs with equivalent sensitivities. This allowed me to explore whether there was any benefit, in terms of diagnostic accuracy, in using model derived probability thresholds to guide management rather than simply using different CA125 cut-offs. I applied the thresholds to the study cohort and prepared schemata to illustrate the potential implications of using different model thresholds in my study cohort.

6.3 Results

6.3.1 Study cohort

After applying the additional exclusion criteria to the baseline cohort, 29,962 women remained and formed the cohort for this study (**Figure 6.1**). Of these, 279 (0.9%) were diagnosed with ovarian cancer in the 12 months following CA125 testing.

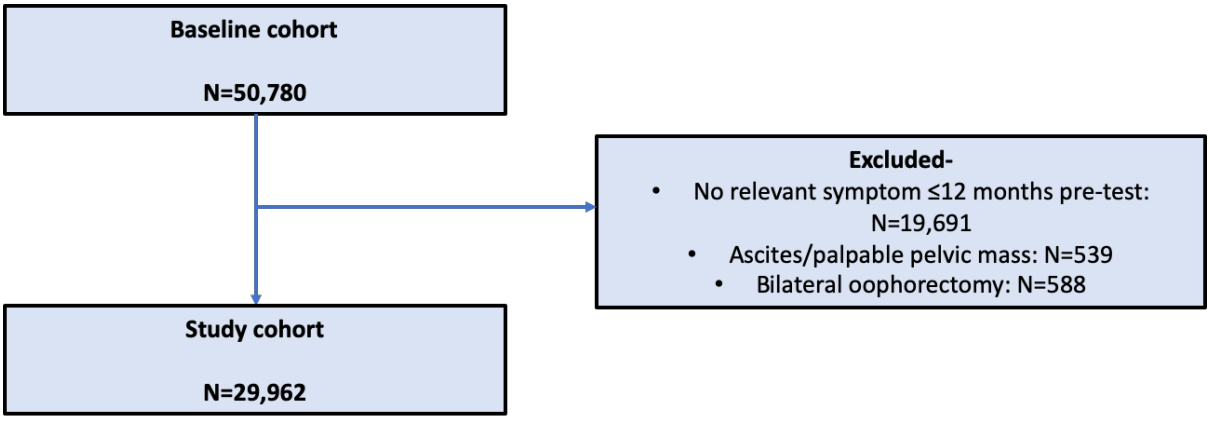


Figure 6.1. Application of additional study criteria to the baseline cohort.

The baseline characteristics of the cohort are shown in **Table 6.4**. The mean patient age was 55 years (range 18-101 years, standard deviation: 15 years). Ethnicity was missing for 1,234 (4.1%), height for 1,721 (5.7%) and BMI for 2,986 (10%) women. 14.2% of women had missing data for at least one of these variables. Where recorded, mean height and BMI were very similar to those reported for women in the Health Survey of England.[292]

Table 6.4. Cohort baseline characteristics.

Figures shown are numbers, (percentage), [standard deviation] and {interquartile range}.

Variable	N=29,962
Risk / protective factors	
Mean age (years)	55 [15]
Ethnicity:	
White	26,511 (88.5)
*Other ethnicities**	2,217 (7.4)
Mean height (cm)	162 [6.8]
Median BMI (Kg/m^2)	25.8 {22.8-29.7}
Personal history breast cancer	1,168 (3.9)
Symptoms	
Abdominal/pelvic pain	17,538 (58.5)
Appetite loss	203 (0.7)
Bloating	5,649 (18.9)
Distension	821 (2.7)
CIBH	5,808 (19.4)
Fatigue	3,968 (13.2)
Urinary frequency/urgency	1,503 (5)
≥50yrs of age with new IBS	286 (1)
Weight loss	1,168 (3.9)
Blood biomarkers	
Median CA125	12 {8-17}
Albumin:	
Not tested	3,723 (12.4)
<35 g/L	834 (2.8)
≥35 g/L	25,405 (84.8)
Haemoglobin:	
Not tested	1,648 (5.5)
<12g/dl	3,089 (10.3)
≥12g/dl	25,225 (84.2)
Platelets:	
Not tested	1,679 (5.6)
$<300x10^9/L$	20,442 (68.2)
$300\text{-}449x10^9/L$	7,314 (24.4)
$≥450x10^9/L$	527 (1.8)
CRP:	
Not tested	13,181 (44)
<3 mg/L	6,907 (23.1)
3-9.99 mg/L	7,370 (24.6)
≥10 mg/L	2,504 (8.4)

The majority of women (n=17,538, 58.5%) had abdominal/pelvic pain recorded in the year prior to CA125 testing. Bloating and change in bowel habit were also common, recorded for 5,649 (18.9%) and 5,808 (19.4%) women respectively. Appetite loss was the least common symptom, recorded for only 203 (0.7%) women.

The majority of patients had blood tests performed in the year before CA125 testing. An albumin level was recorded for 26,239 (87.6%), haemoglobin level for 28,314 (94.5%), platelets for 28,283 (94.4%) and CRP for 16,781 (66%) women.

6.3.2 Final models

Of the 19 candidate variables considered for inclusion in Model 2, age, ethnicity, BMI, height, abdominal/pelvic pain, distension, CA125 level, platelet level and albumin level were retained after backward elimination procedures. The coefficients and odds ratios for variables included in the models are presented in **Table 6.5** for reference. In Model 2, the Other ethnic category was associated with lower odds of ovarian cancer than White ethnic category. Increasing log BMI and height were both associated with increasing odds of ovarian cancer, as were the coding of abdominal/pelvic pain and abdominal distension. Unexpectedly, the odds of ovarian cancer were lower in groups of women with a recorded platelet count and a recorded albumin level than in the no-test group, although confidence intervals spanned 1 for several of these test groups.

Table 6.5. Coefficients and odds ratios for variables included in Model 1 and Model 2.

	Model 1		Model 2	
Variable	**Coef. (95% CI)**	**Odds ratio (95% CI)**	**Coef. (95% CI)**	**Odds ratio (95% CI)**
Baseline risk factors				
Age:				
Age spline 1	-0.079 (-0.136 – -0.022)	0.924 (0.873 – 0.978)	-0.077 (-0.136 – -0.018)	0.926 (0.873 – 0.982)
Age spline 2	0.537 (0.223 – 0.852)	1.712 (1.250 – 2.345)	0.520 (0.199 – 0.841)	1.682 (1.220 – 2.319)
Age spline 3	-2.169 (-3.910 – -0.428)	0.114 (0.020 – 0.652)	-2.025 (-3.798 – -0.253)	0.132 (0.022 – 0.777)
Age spline 4	1.712 (-0.516 – 3.940)	5.539 (0.597 – 51.420)	1.542 (-0.724 – 3.808)	4.674 ($2.81x10^{-6}$ – 0.018)
Ethnicity:				
White			Reference	Reference
Other			-0.906 (-1.756 – -0.055)	0.404 (0.173 – 0.947)
Log BMI			0.965 (0.224 – 1.705)	2.624 (1.251 – 5.503)
Height (cm)			0.040 (0.017 – 0.062)	1.041 (1.017 – 1.064)
Symptoms				
Abdominal / pelvic pain			0.412 (0.089 – 0.735)	1.510 (1.093 – 2.087)
Distension			0.648 (0.034 – 1.261)	1.911 (1.035 – 3.530)
Tests				
Log CA125:				
Log CA125 spline 1	1.129 (-2.386 – 4.643)	3.092 (0.092 – 103.862)	1.043 (-2.429 – 4.516)	2.839 (0.088 – 91.447)
Log CA125 spline 2	-7.114 (-26.469 – 12.241)	0.0008 ($3.19x10^{-12}$ – $2.07x10^{5}$)	-6.592 (-25.805 – 12.622)	0.001 ($6.21x10^{-12}$ – $3.03x10^{5}$)
Log CA125 spline 3	82.537 (-17.899 – 182.973)	$7.01x10^{35}$ ($1.69x10^{-08}$ – $2.91x10^{79}$)	78.551 (-21.420 – 178.521)	$1.30x10^{34}$ ($4.98x10^{-10}$ – $3.39x10^{77}$)

Log CA125 spline 4	-143.749 (-267.564 – -19.934)	$3.72x10^{-63}$ ($6.3x10^{-117}$ – $2.20x10^{-09}$)	-137.307 (-260.776 – -13.839)	$2.33x10^{-60}$ ($5.6x10^{-114}$ – $9.77x10^{-07}$)
Platelets: No test			Reference	Reference
<300x10^9/L			-0.699 (-1.350 – -0.048)	0.497 (0.259 – 0.953)
300 – 449 x10^9/L			-0.378 (-1.053 – 0.297)	0.685 (0.349 – 1.346)
≥450 x10^9/L			-0.103 (-0.885 – 0.678)	0.902 (0.413 – 1.971)
Albumin: No test			Reference	Reference
<35 g/L			-1.241 (-1.951 – -0.531)	0.289 (0.142 – 0.588)
≥35 g/L			-0.106 (-0.625 – 0.413)	0.899 (0.535 – 1.511)

Coef. = variable coefficient. CI = confidence interval

6.3.3 Discrimination and validation

The AUC of CA125 alone, calculated directly from the study dataset, was 0.932. Model 1 and Model 2 both had an AUC of 0.938, when calculated directly from the dataset (apparent performance) (**Table 6.6**). On cross-validation, both models had an AUC of 0.935. There was little difference between apparent and cross-validation AUCs for each model, indicating that overfitting/optimism was minimal. A ROC curve for CA125 and cross-validation ROC curves for the models are included in **Appendix J** for reference. Both models had calibration slopes close to 1, indicating that overall they did not markedly over or underestimate risk. However, confidence intervals were wide.

Table 6.6. Model discrimination and calibration.

Model	Apparent AUC	Cross-validation AUC	Cross-validation calibration slope (95% CI)
Model 1	0.938	0.935	1.01 (0.606-1.42)
Model 2	0.938	0.935	1.05 (0.673-1.42)

AUC = area under the curve

6.3.4 Thresholds for further investigation

As the more parsimonious Model 1 exhibited the same AUC and similar calibration metrics to Model 2, my evaluation of thresholds for further investigation focussed on Model 1. The diagnostic accuracies of Model 1 thresholds (for the detection of ovarian cancer within 12 months of CA125 testing) are shown in **Table 6.7**. In order to examine whether using the Model had any benefit over using different CA125 cut-offs, these are presented alongside CA125 cut-offs with equivalent sensitives. At the ≥1% probability threshold, the specificity of Model 1 was 3.1% higher than a CA125 cut-off with the same sensitivity (≥23 U/ml), while there was a less marked difference at higher model probability thresholds. For all model thresholds, the PPV was higher than for CA125 cut-offs with equivalent sensitivities.

The ≥1% model threshold was 7.9% more sensitive, but 5.4% less specific, than the national CA125 cut-off. The ≥2% model threshold had the same sensitivity as the national CA125 cut-off and similar specificity. A ≥3% model threshold was 2.9% less sensitive and 2.4% more specific than the national CA125 cut-off.

Table 6.7. Diagnostic accuracy metrics for a range of Model 1 thresholds and CA125 cut-offs with equivalent sensitivities.

	Sensitivity	Specificity	PPV	NPV
≥1% model probability	86.4 (81.8-90.2)	89.1 (88.8-89.5)	6.9 (6.1-7.8)	99.9 (99.8-99.9)
CA125 of ≥23 U/ml	86.4 (81.8-90.2)	86.0 (85.6-86.4)	5.5 (4.8-6.2)	99.9 (99.8-99.9)
≥2% model probability	78.5 (73.2-83.2)	94.7 (94.5-95.0)	12.2 (10.8-13.9)	99.8 (99.7-99.8)
Ca125 of ≥35 U/ml	78.5 (73.2-83.2)	94.5 (94.3-94.8)	11.9 (10.4-13.4)	99.8 (99.7-99.8)
≥3% model probability	75.6 (70.2-80.5)	96.9 (96.7-97.1)	18.5 (16.3-20.9)	99.8 (99.7-99.8)
CA125 of ≥39 U/ml	75.6 (70.2-80.5)	95.6 (95.3-95.8)	13.8 (12.1-15.7)	99.8 (99.7-99.8)

Figure 6.2 illustrates the potential implications of applying the different model thresholds to the study cohort of 29,962 women. In comparison to the national CA125 cut-off, applying a probability threshold of ≥1% resulted in an additional 1,622 women being identified for further evaluation for ovarian cancer. Of these, 22 (1.4%) had ovarian cancer (i.e. an additional 1 in every 74 women identified for further evaluation had ovarian cancer). Applying a ≥3% model probability threshold, instead of the current CA125 cut-off, resulted in 706 fewer women being identified for further evaluation of whom 8 (1.1%) had ovarian cancer. Applying a ≥2% model threshold, instead of the current CA125 threshold, resulted in 58 fewer women being identified for further evaluation of whom none had ovarian cancer.

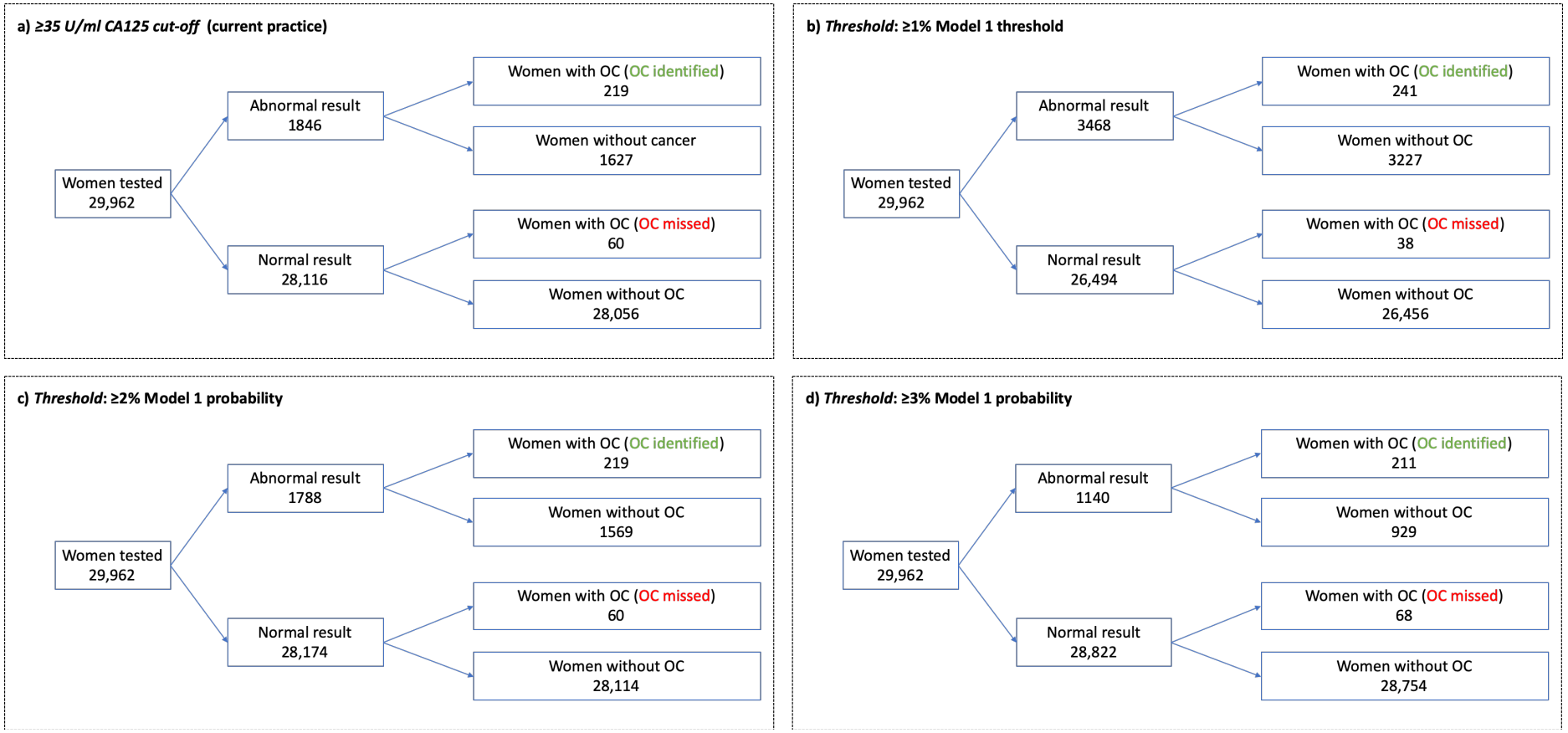


Figure 6.2. Implications of applying CA125 and model thresholds to the study cohort.

6.4 Discussion

6.4.1 Summary

A model consisting of CA125 and age demonstrated excellent discrimination and calibration for the identification of ovarian cancer in symptomatic women undergoing CA125 testing in primary care in England. Incorporating additional baseline risk factors, symptom type and routine blood test results within the model did not improve performance. Although the AUC of this model was only slightly higher than that of CA125 alone, at a fixed sensitivity Model 1 showed superior specificity and PPV at a range of thresholds. When a ≥1% probability threshold was applied to my cohort, rather than the current CA125 cut-off (≥35 U/ml), one in 74 of the additional women identified by the model had ovarian cancer. Key study findings are summarised against objectives in **Box 6.2**.

viii) To develop and internally validate an ovarian cancer diagnostic prediction model incorporating symptoms, test results (including CA125), and risk factors

- A diagnostic prediction model (Model 2) was developed which incorporated nine variables:
  - *Baseline risk factors*: age, ethnicity, BMI, height
  - *Symptoms*: abdominal/pelvic pain, distension
  - *Test results*: albumin level, platelet count and CA125 level
- This model showed excellent discrimination, with minimal evidence of optimism on cross-validation
- However, this performance was matched by a much simpler model comprising age and CA125 level (Model 1)

ix) To explore the potential diagnostic implications of implementing 'action thresholds', based on prediction model derived estimated ovarian cancer probabilities, within primary care

- When applied to my study cohort of 29,962 women:
  - A ≥3% Model 1 probability threshold was 2.9% less sensitive, 2.4% more specific and had a 6.6% higher PPV than the current CA125 cut-off
  - A ≥1% Model 1 probability threshold was 7.9% more sensitive but 5.6% less specific, with a 5% lower PPV, than the current CA125 cut-off
- On applying a ≥1% Model 1 probability threshold to my cohort, rather than the current CA125 cut-off, one in 74 of the additional women identified by the model had ovarian cancer

Box 6.2. Key results against chapter objectives.

6.4.2 Study limitations

The dataset used in this study provided information on a wide range of variables for a large number of women. However, use of routinely collected data did limit the candidate variables I considered for Model 2. For example, I did not include family history of ovarian cancer, a

strong risk factor, due to concerns over data completeness and recording bias. Concerns over the lack of coding of symptoms within CPRD led me to exclude 19,691 women (who did not have a symptom coded in the 12 months prior to CA125 testing) from my study cohort. In so doing, my sample size was reduced so the risk of model overfitting and of imprecise probability estimates increased. This limited the number of candidate variables I included in model development. While this approach ensured that I developed my models in a population of women with at least one relevant symptom of possible ovarian cancer, it did not entirely resolve the problem with missing symptom information. For example, a woman may have abdominal pain coded, and so be included in the study cohort, but a second symptom e.g. bloating, may be recorded in free text rather than being coded. There is also the possibility that the models may not be fully generalisable to women with a symptom which has not been coded, as symptom coding may be a marker for symptom/disease severity or GP concern. An alternative approach would have been to include all women from my baseline cohort in this study, regardless of whether they had a coded symptom or not. With such an approach I would have been quantifying the predictive ability of having a symptom *coded* (acknowledging that some patients may have un-coded symptoms). This would have been appropriate given the way in which clinical prediction models are often used in clinical practice: an algorithm embedded within the GP software searches the patient record and provides the clinician with a risk level and clinical recommendation based on the *coded* data.

I was conscious of the of EPV recommendations when selecting candidate variables for model development. The final EPV of 9 for Model 2 was just below the commonly recommended EPV of 10. However, following model development I performed a post hoc calculation of the minimum recommended sample size for this study, in line with the new guidelines from Riley *et al* published in 2020, using the *pmsamplesize* command in Stata.[293] Given an outcome proportion of 0.093, an EPV of 32 and a Cox-Snell R^2 (a measure of model performance) of 0.013,[§§] the minimum recommended sample size for this study was 21,994 women. This provides added reassurance that my sample size was adequate.

[§§] The R^2 was calculated using a predeveloped excel template (provided by Prof Gary Abel) and assumed an AUC of 0.92.

Although potential predictors of ovarian cancer were carefully identified from the literature in order to develop Model 2, there are several possible explanations as to why their inclusion did not ultimately improve model performance.

In contrast to the prediction model studies reviewed in **Chapter 6**, all women in my model development cohort had at least one symptom of possible ovarian cancer. Therefore, the predictive effect of having a relevant symptom was, to some extent, already accounted for prior to model development. In addition, I only included women who had been selected by their GP for CA125 testing. GPs do not test all women with symptoms such as bloating or abdominal pain for ovarian cancer: my cohort is likely to consist of those women who, on the basis of their clinical presentation, were judged at to be at higher risk. It is possible that if my models were developed as part of a prospective study in which all women with relevant symptoms presenting to primary care were tested for CA125, the risk associated with having a symptom would have been more accurately accounted for and Model 2 may have outperformed Model 1. However, the aim of this study was not to develop a model for use in *all* women presenting to their GP with possible symptoms of ovarian cancer, but to develop a model to aid the triage of women who *are* tested for CA125 by their GP.

Missing data on symptoms may have limited the performance of Model 2 as the predictive effect of a given symptom could not be fully accounted for. I only included women with a coded symptom in my model development cohort, but symptom coding may indicate more severe symptomology or a greater suspicion of ovarian cancer,[187] so my model development cohort may have had a higher overall risk of ovarian cancer than my baseline cohort. It is possible that if Model 2 had been developed using the entire CA125 tested cohort, and if all symptoms were accurately coded, strongly predictive symptoms such as abdominal distension may have had a greater predictive effect within the model. Even within my model development cohort, missing data on symptoms could have limited model performance as although all patients had one symptom coded they may have had other symptoms which were not coded and so the predictive effect of a given symptom may not have been accurately accounted for within the model.

I selected baseline risk factors as candidate variables for Model 2 based on the published literature. However, the risk associated with most of these variables in the literature was relatively modest e.g. an increase in height of 10cm was associated with a relative risk of 1.16 in one meta-analysis.[36] While both height and BMI were retained within Model 2, their contribution to ovarian cancer prediction within a symptomatic population, when included alongside CA125 (a very strong predictor), was ultimately negligible.

I internally validated models using a cross-validation approach. This indicated that there was minimal model overfitting (which can be caused by limited sample size and data driven variable selection techniques). However, internal validation does not provide an indication of the generalisability of the models outside the study sample. This study utilised the CPRD GOLD database, which collects data from GP practices using Vision software. The structure and recording of variables in other databases, which collect primary care data from practices using alternative IT software (such as QResearch which collects data from GP practices using EMIS software), may differ from that of CPRD GOLD. In addition to structural differences there are also likely to be differences in the populations and cancer incidence rates between databases e.g. due to different UK regional coverage. For these reasons, models tend to perform less well on external validation. If the models developed in this study are to be implemented within English primary care, they should ideally be externally validated to ensure adequate performance.

6.4.3 Comparison with existing literature

In the systematic review presented in **Chapter 5** I did not identify any primary care diagnostic prediction models which included symptoms and CA125. Existing primary care diagnostic prediction models, such as QCancer Ovarian,[54] were developed in general primary care populations (which included both women *with* and *without* symptoms) with the aim of identifying higher risk women for tests such as CA125, whereas I developed the models in this study within an entirely symptomatic CA125 tested population. Similarly, the performance metrics of secondary care diagnostic models, such as the Risk of Ovarian Malignancy Algorithm (ROMA) and the Risk of Malignancy Index (RMI), are not comparable as they were developed in women known to have a pelvic mass, with the aim of distinguishing between benign and malignant masses.[294,295] These models also include more specialist tests (ROMA:

the HE4 biomarker, RMI: transvaginal ultrasound). This is the first study to develop and internally validate multivariable CA125 predicated models within the symptomatic primary care population.

6.4.4 Model variables

Baseline patient risk factors, routine blood test results and the type of symptoms with which patients present, have previously been found to be predictive of ovarian cancer in general practice,[296] and so were included as candidate variables in this study. Variables in each of these categories were predictive of ovarian cancer (to some extent) and were therefore selected during data driven procedures for inclusion in Model 2. However, the odds ratios for the majority of these variables were modest and confidence intervals wide. The numbers of patients with some model variables were low e.g. only 2.7% of women had distension, which limits the contribution a variable can have to the overall AUC of a model. Surprisingly, the association of platelet count and albumin level with ovarian cancer within Model 2 were the opposite of what was anticipated - odds were lower in women who had a test than in those who had not had a test. The reasons for this are unclear, but it is possible that GPs request multiple blood tests, including CA125, if they are uncertain of the cause of a symptom, whereas they request CA125 in isolation if they have a high suspicion of ovarian cancer.

I did not develop a prediction model containing only CA125 level. Age is available for all patients within GP records and laboratory systems and implementing a model with age and CA125 in clinical practice would be no more complex than implementing one consisting solely of CA125 level. By contrast, the number of variables in Model 2 would make its implementation more complex and its use more challenging (e.g. a user would have to enter BMI, height and ethnicity manually if not coded in the patient record) for no material gain over Model 1 in terms of diagnostic performance.

6.4.5 Clinical relevance of findings

If alternative 'action thresholds' were to be implemented in clinical practice, in place of the ≥35 U/ml CA125 cut-off, the results of this study indicate that there is diagnostic benefit in selecting them using Model 1 rather than simple CA125 cut-offs, particularly at lower risk levels. The model has the additional benefit that the individual probability of ovarian cancer

can be provided to the patient, if they wish it, and be used to help inform individual decisions on whether to have further investigation or not. It also allows action to be taken at particular risk thresholds set out in guidelines e.g. the ≥3% NICE threshold for urgent investigation. Use of a low model probability threshold in clinical practice e.g. ≥1%, in place of the ≥35 U/ml CA125 cut-off, might expedite diagnosis of some ovarian cancer cases but would increase the number of women without cancer who undergo unnecessary further investigation. In **Chapter 7**, I consider this in more detail and discuss how Model 1 thresholds might be incorporated within the ovarian cancer diagnostic pathway in England.

6.5 Conclusion

A model consisting of age and CA125 level performed well for the detection of ovarian cancer in symptomatic women in English primary care. Including additional variables within the model did not improve performance. Following external validation, this model could be used to help select women for further investigation based on individual cancer probability.

Chapter 7. Discussion

7.1 Summary

The overarching aim of this thesis was to evaluate the diagnostic performance of CA125 for the detection of cancer when used in primary care and to develop and evaluate novel approaches to improve its performance and clinical utility.

In **Chapter 3,** I presented the first large study to evaluate the accuracy of CA125 in primary care, which demonstrated that the test performs well for the detection of ovarian cancer in this setting. At 10.1%, the PPV of CA125 at the national cut-off (≥35 U/ml) was 12 times higher than the estimate NICE used to inform the development of guidelines covering testing for ovarian cancer in primary care in 2011.[114]

Through logistic regression analyses, I demonstrated that the probability of ovarian cancer varies markedly depending on the specific CA125 level and patient age. This means that the overall PPV of CA125 at the conventional cut-off is of limited use to clinicians and patients when interpreting their individual CA125 results. Instead, I present (and have made freely available)[209] estimated ovarian cancer probabilities (based on CA125 level and age) which could be used by patients and GPs to inform individual decisions about the need for further investigation. This information could also be used by policy makers to inform guidelines so that thresholds for referral are based on an individual's cancer probability rather than a generic CA125 cut-off.

This thesis primarily focusses on ovarian cancer detection. However, perhaps the most striking finding was that older women without ovarian cancer but with elevated CA125 levels had a very high probability of having another form of cancer (e.g. 20.4% if ≥50 years with a CA125 ≥35 U/ml). This reflects the non-specific nature of ovarian cancer symptoms (which overlap with symptoms of other malignancies) and the non-specific nature of CA125 itself (which can be raised in a variety of other malignancies). Most of these other cancers have more accurate triage tests, so CA125 is unlikely to be of value as an investigation for any specific cancer (other than ovarian) in primary care. However, recognition by clinicians,

patients and policy makers that a high CA125 level is a multi-cancer risk marker in older women, could aid earlier diagnosis.

In **Chapter 4**, I explored the association between pre-diagnostic, primary care CA125 test result ('normal' vs 'abnormal') with test-diagnosis interval and stage in women with ovarian cancer. I found that, although women with normal CA125 results took markedly longer to be diagnosed than those with abnormal CA125 results, they commonly had indolent tumour types and were usually diagnosed at an early stage. This should provide some reassurance for those using, and being tested for, CA125 in primary care. However, given the study design, I was unable to determine what proportion of women with false negative CA125 results progressed during their extended test-diagnosis intervals. Even progression within the early stage cancer category (e.g. from I to II), or an increase in tumour burden, will have prognostic implications for women. Given the long test-diagnostic intervals experienced by a proportion of women with normal CA125 results (>6 months in 15%), their expedited diagnosis could improve outcomes.

In **Chapter 5**, I presented a systematic review of symptom predicated multivariable tools for the detection of ovarian cancer. I identified 21 tools, including several secondary care checklists which incorporated CA125 alongside symptoms, but no CA125 containing prediction models. Four tools (none of which incorporated CA125) had undergone external validation and exhibited similar (moderate) diagnostic accuracy. Such tools could be used in primary care to select women for ovarian cancer tests such as CA125, but further evaluation is needed to determine their acceptability, cost-effectiveness and clinical impact. The variables included in tools fell into three categories: 1) symptoms, 2) risk / protective factors and 3) tests. This helped guide my selection of candidate variables for my subsequent prediction modelling study (**Chapter 6**).

In **Chapter 6**, I described the development and internal validation of ovarian cancer diagnostic prediction models for use in symptomatic women who have had a CA125 test performed in primary care. In women with a relevant symptom code in their GP records in the year prior to CA125 testing (a sub-sample of my baseline cohort), a model consisting of age in years and CA125 level exhibited excellent discrimination on cross-validation (AUC: 0.94%). Including

information on symptom type, baseline risk factors and routine blood test results in the model, alongside CA125 and age, did not improve performance.

I explored the sensitivity and specificity of various model-based probability thresholds and considered the implications of applying them to the study cohort. A model threshold, set to detect women with a ≥1% estimated probability of ovarian cancer, had a 7.9% higher sensitivity but 5.4% lower specificity compared with a CA125 cut-off of ≥35 U/ml. The overall PPV of this ≥1% model probability threshold remained high at 6.9%. Applying this threshold to my cohort in place of a CA125 of ≥35 U/ml, 1 in every 74 additional women identified had ovarian cancer. Using this model threshold in clinical practice as an action threshold might help identify more women with ovarian cancer in a timely way, but would also result in more women without ovarian cancer undergoing unnecessary further investigation.

Collectively, the evidence presented in this thesis indicates that CA125 is a useful test for the detection of ovarian cancer in primary care and that a high CA125 level should also raise the suspicion of other types of cancer, especially if ovarian cancer has been excluded. A relatively simple multivariable prediction model (age + CA125) can provide patients and clinicians with the estimated probability of ovarian cancer, which could be used to inform individual decisions about the need for further investigation. This information could also be used by policy makers, such as NICE, when developing guidelines so that referral recommendations are based on cancer probability.

7.2 Strengths and limitations

I have set out the strengths and limitations of each study in detail within the relevant chapters. Here, I summarise important strengths and limitations which should be considered when interpreting my thesis findings and their clinical implications.

I consider the use of data from the CPRD as a major strength. The CPRD, which is considered generally representative of the UK population in terms of key demographics,[178] provided information on 50,780 CA125 tested women, making this one of the largest studies ever

conducted on CA125 or the diagnosis of ovarian cancer to date.[***] Moreover, by using routinely collected data, I could evaluate the performance of the test as it is used within routine general practice. Blood test results are automatically transferred from laboratories to GP records (and then collected within the CPRD). While I identified some issues with CA125 recording (e.g. issues with units), the data generally appeared to be of high quality. Another strength of this study was the use of NCRAS data to identify outcomes. Not only is NCRAS considered more accurate than CPRD and HES in terms of cancer recording,[198] but it also provides information on tumour type and stage at diagnosis which were used within this thesis.

However, routinely collected data has its limitations. A significant limitation of CPRD and NCRAS is the presence of missing data. As symptoms are not always coded in GP records, it was unclear why some CA125 tests were performed in the baseline cohort. The recording of stage has improved in NCRAS in recent years but is not complete:[198] stage was missing for 16% of ovarian cancers in my baseline cohort. Moreover, missing stage was more common for women with normal CA125 levels prior to diagnosis (likely due to the higher proportion of borderline tumours in this group), which could have introduced bias into my stage analysis in **Chapter 4**. Due to concerns over the coding of symptoms, I excluded 39% of my baseline cohort (who did not have a relevant symptom code in the year prior to CA125 testing) from my model development cohort in **Chapter 6**. Excluding women without a symptom code may have introduced selection bias, as symptoms are more likely to be coded (rather than recorded in the free text) in patients with, than without, cancer.[187] However, a model consisting CA125 and age (analogous to Model 1) was also developed using the whole baseline cohort (**Chapter 3)**.

Cross-validation of the CA125 level and age-based model in **Chapter 6** indicated that there was minimal overfitting. Nevertheless, it is possible that, due to differences in populations or CA125 testing patterns or recording, the model may perform differently when evaluated in alternative databases such as QResearch or CPRD AURUM,[179] or in clinical practice using

[***] In the UKCTOCS trial slightly fewer women (50,640) had a CA125 test.[154] The work presented in **Chapter 3** includes more CA125 tested women than any other study at the time of writing (April 2021).

different GP records systems. TRIPOD guidelines and Steyerberg recommend that external validation should ideally be performed prior to clinical implementation of new prediction models to ensure generalisability of predictions.[162,171] This was not feasible as part of this thesis, but I am planning to externally validate my CA125 and age based model as part of a post-doctoral study.

7.3 Key clinical implications

I have presented the clinical implications of each research study in detail in the relevant chapters. Here, I consider how the findings of this thesis as a whole could be used to improve cancer diagnostic pathways.

7.3.1 Ovarian cancer

Current NICE guidelines on ovarian cancer detection were developed using an estimated PPV for CA125 which, although it was based on the best available evidence at the time, this thesis has shown to be incorrect.[114] Economic modelling and the recommendation for sequential primary care testing (CA125 then ultrasound) before referral, was predicated on this estimate. In light of the findings presented in this thesis I recommend that these guidelines be reviewed and a fresh economic evaluation performed.

My logistic regression analyses indicated that older women with CA125 levels above 35 U/ml already have a probability of ovarian cancer ≥3% (e.g. 32 U/ml equates to a 3% probability in a 70-year-old). Yet, under NICE guidelines, these women are required to have an ultrasound in primary care before they qualify for an urgent cancer referral. As primary care ultrasounds usually take several weeks to be performed in England,[297] this could delay ovarian cancer diagnosis. In addition, NICE requires that *both* the primary care CA125 test and ultrasound be abnormal for a woman to qualify for an urgent specialist referral. Given that ultrasound can be normal in women with elevated CA125 levels prior to ovarian cancer diagnosis,[152] such a sequential primary care testing approach also risks delaying diagnosis due to false negative ultrasound results. The models presented in this thesis could be used to select women for urgent cancer pathway referral, in line with the NICE 3% threshold, thereby helping to ensure that women at high risk of undiagnosed cancer receive prompt specialist assessment, diagnosis and treatment.

In addition to selecting women at high risk for urgent referral, the CA125 and age-based model presented in this thesis could be used to identify women at 'low risk but not no risk' of ovarian cancer. For example, those with a probability of ovarian cancer between 1-3%, who might be offered non-urgent evaluation or interval re-assessment. This evaluation could involve interval CA125 re-testing, as most women who are retested in primary care prior to diagnosis have rising CA125 levels (**Chapter 4**), a referral for a non-urgent transvaginal ultrasound or, given the additional risk of non-ovarian cancers, perhaps a referral to a Rapid Diagnostic Centre (RDC) (**Section 7.3.4**). As demonstrated in **Chapter 6**, applying a 1% model threshold will identify more ovarian cancers than the current CA125 cut-off. This could help reduce diagnostic delay related to false negative CA125 results (**Chapter 4**), which has the potential to improve patient outcomes.

Figure 7.1 illustrates an example of a two-tier risk-stratified triage approach using model probability thresholds. Such an approach is likely to result in more non-urgent investigation in primary care ('low risk but not no risk' women) and more urgent cancer referrals (higher risk women). Any such change in guidelines would require a full health economic evaluation, to assess the potential impact of the strategy on the healthcare service and on patients, as most of those investigated would ultimately not be diagnosed with ovarian cancer. A health economic evaluation would also be valuable in ensuring that the most appropriate model thresholds are chosen for use in the risk-based triage approach. The model thresholds presented within this thesis are of particular relevance to the healthcare system in England. However, following validation in appropriate local data-sets, the model could be used to select women for further investigation for ovarian cancer in line with any regional or national threshold.

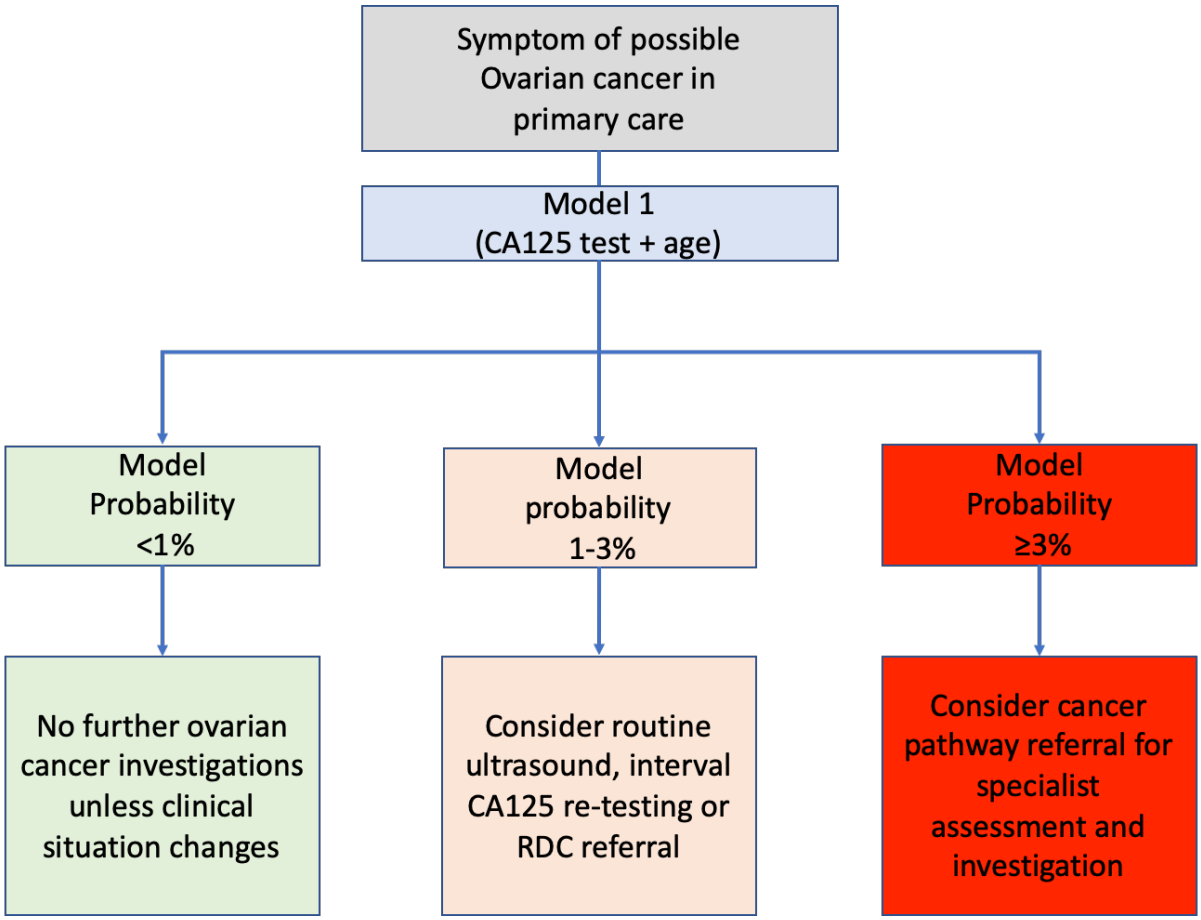


Figure 7.1. Example risk-based triage system employing model probability thresholds.

7.3.2 CA125 testing in young women

Given that ovarian cancer is rare in younger women, it was surprising that 39% of CA125 tests in my baseline cohort were performed in women under 50 years of age. Age is known to play a part in whether women opt for cancer investigation, with a large vignette study finding that women aged 40-59 are more likely to opt for investigation when presented with the same clinical scenario as women aged ≥70 years.[200] While further research would be needed to understand why there is such a high rate of CA125 testing in younger women, it represents relative over-testing in younger (lower risk) women compared to older (higher risk) women. The models presented in this thesis take account of age and, if used in practice, should aid the interpretation of CA125 results in younger women, who have relatively low ovarian cancer probabilities even at CA125 levels around the national cut-off. This could help avoid unnecessary further investigation. Reducing the number of younger women who undergo inappropriate CA125 tests in the first place could also help prevent unnecessary further investigations and referrals. Guidelines and GP education material should reinforce that

ovarian cancer is relatively rare in younger women and also that CA125 is less accurate (lower AUC and lower sensitivity, specificity and PPV at the ≥35 U/ml cut-off) in this age group.

7.3.3 Other cancers

A key clinical message of this thesis is that a high CA125 level in older women in primary care should raise the suspicion of non-ovarian cancer if ovarian cancer has been excluded. International guidelines, including those produced by NICE, should be updated to highlight this. Various approaches could be taken to help identify these cancers in a timely way. These are considered in the next section.

7.3.4 The diagnostic pathway

This thesis focusses on the initial testing step within the diagnostic pathways for ovarian cancer, but timely detection of ovarian cancer also depends on the accuracy of subsequent steps of the pathway. The refining ovarian cancer test accuracy scores (ROCkeTS) study, a large prospective study evaluating a range of diagnostic tests and algorithms for ovarian cancer in UK secondary care (including different approaches for ultrasound interpretation), is due to report later in 2021. ROCkeTS could provide insight into the most appropriate post-CA125 testing strategy for ovarian cancer.[298] However, the research presented in this chapter highlights the need for any such strategy not only to consider ovarian cancer, but also to consider non-ovarian cancers. Transvaginal ultrasound may detect endometrial cancer and some cancers which cause marked ascites, but is likely to miss many (if not most) CA125 elevating non-ovarian malignancies in addition to a proportion of ovarian cancer. Further research is required to identify the most appropriate diagnostic strategy, but I would like to highlight three possible approaches to post-CA125 testing that I believe warrant consideration and further evaluation (**Figure 7.2**).

a) ***Increased awareness and targeted cancer investigation***

NICE guidelines recommend that if a woman has an abnormal CA125 test but a normal ultrasound scan, the GP should consider alternative causes for the patient's symptoms.[57] If GPs are made aware of the high risk of non-ovarian cancer in these women they could use available patient information (including patient risk factors and symptoms) and clinical judgement to select the most appropriate targeted cancer investigations or make a specialist

referral where appropriate. This could be supported by education programs and by updating NICE guidelines to highlight the high cancer risk in this group and to encourage GPs and gynaecologists to consider whether further cancer specific investigations are needed.

b) ***Referral to a Rapid Diagnostic Centre (RDC)***

The symptoms with which patients present to their GP prior to cancer diagnosis are often non-specific and, as discussed in **Chapter 3**, may be caused by multiple different types of cancer.[57,299] A recent study from Denmark found that in the year prior to diagnosis of an abdominal cancer, patients frequently underwent specific investigations more appropriate for the diagnosis of another type of cancer.[300] For example, >10% of patients diagnosed with oesophageal cancer had an abdominal ultrasound scan and >7% of patients with renal cancer had a gastroscopy or colonoscopy procedure. In patients with non-specific symptoms, the risk of individual cancers may be relatively low, but the combined risk of having some form of cancer can be relatively high.[57,301] This reflects the situation in women 50 years or older with high CA125 levels but no ovarian cancer.

Multidisciplinary Diagnostic Centres (MDCs) were first set up in Denmark in 2012,[302] and have now been successfully trialled across five sites in England as part of the Accelerate Coordinate Evaluate (ACE) program.[303] Their aim is to expedite diagnosis in patients with non-specific ('low risk but not no risk') symptoms by providing rapid assessment and access to a range of urgent cancer tests which are often not available to GPs in the community (e.g. urgent abdominopelvic CT). Based on this model, a roll-out of Rapid Diagnostic Centres (RDCs) across England was recommended as part of the NHS long term plan published in 2019.[68] Specific guidance was also published by NHS England in 2019 which recommended that all Cancer Alliances set up at least one RDC before the end of 2020,[304] but this roll-out has been affected by the COVID-19 pandemic. Women with high CA125 levels (or a high cancer probability derived from the all-cancer model, **Chapter 3**) but a normal ultrasound scan, could be referred directly to RDCs in order to expedite cancer investigation, diagnosis and treatment.

c) ***Direct access urgent CT scans***

CT can detect multiple CA125-elevating cancer types including ovarian, lung and pancreatic cancer.[305–308] It is already recommended in some countries as an alternative (or additional)

investigation to ultrasound in women with symptoms of possible ovarian cancer.[116] CT could serve as a multi-cancer test in women in England with high CA125 levels (or a high all-cancer probability derived from the cancer model, **Chapter 3**) but normal ultrasound scans.[†††]

In his 2020 report on diagnostic services in England, Professor Sir Mike Richards recommended the establishment of 'community diagnostic hubs' which would provide access to a range of specialist investigations (including CT) to GPs.[309] If established, these hubs could facilitate direct GP access to CT investigation in women with high CA125 levels and normal ultrasound scans. Alternatively, CT imaging could be requested in secondary care, in women referred to gynaecology due to suspicion of ovarian cancer, following specialist assessment and investigation for ovarian pathology.

A CT based testing approach would require substantial research prior to any widespread implementation – I discuss this further in **Section 7.4.3.**

[†††] I have already been contacted by a clinical group in England who are considering how best to implement such an approach in their region.

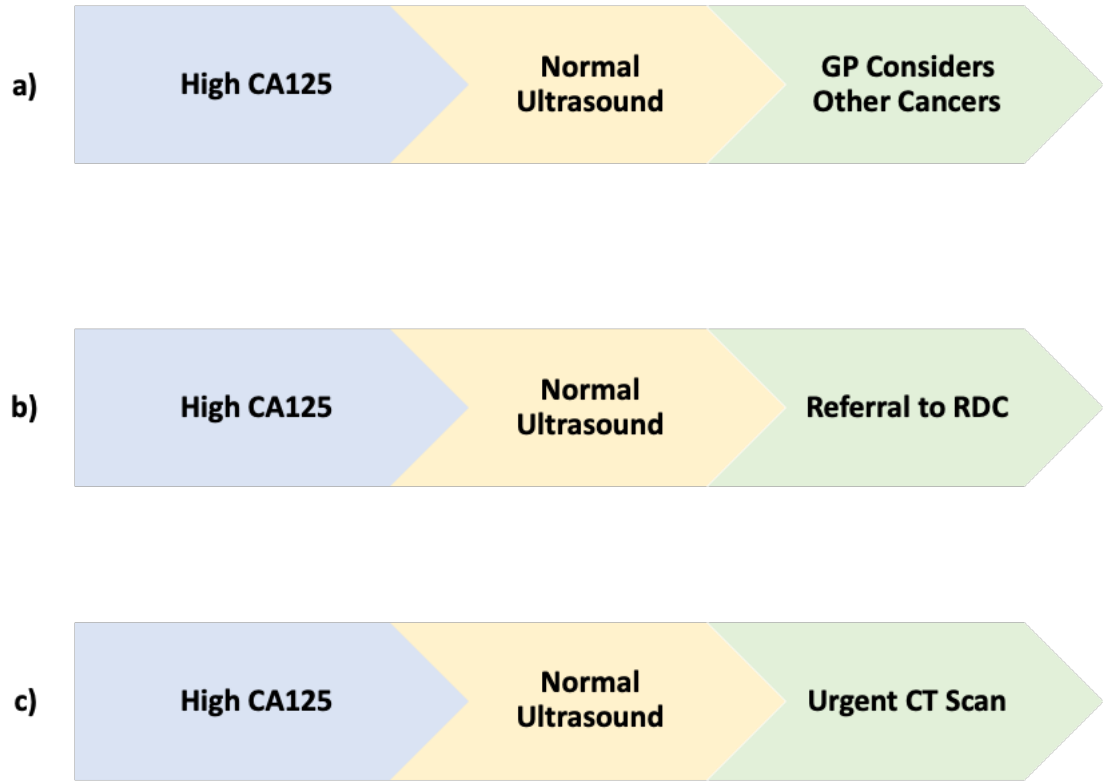


Figure 7.2. Possible cancer diagnostic strategies in women with high CA125 levels but normal ultrasound scans.

7.4 Further research

7.4.1 Ovarian cancer models

While a CA125 and age based diagnostic prediction model has potential clinical utility, significant further work is required to translate this model into clinical practice and to ensure its effectiveness (**Figure 7.3**).

Ideally, the model should undergo external validation in a distinct dataset to ensure generalisability. A number of datasets are available including the large CPRD AURUM database.[179] A health economic evaluation should also be performed to help determine whether a two-tier, risk-based triage approach (discussed above) is likely to be cost effective and to help determine which model probability thresholds should be implemented.

The model must then be integrated within clinical IT systems in a way that is useful for GPs. Studies have reported regulatory and technical challenges in implementing models and this is further complicated by the fact that there are multiple distinct clinical IT systems used by

GPs in the UK and that each system operates in a different way.[310] GPs have reported a number of barriers to the use of eCDS tools, including lack of time (e.g. when data has to be entered manually), prompt fatigue (when the clinical system provides an 'alert' at a given model risk level), interoperability issues with software (when the eCDS tool is not embedded within the GP IT system), lack of trust in the tool and a belief that it does not aid clinical decision making.[259,310] A key advantage of my ovarian cancer model is its simplicity: only age and the CA125 level are required. This means that it could either be integrated within laboratory IT systems or GP software. The probability estimate and a clinical recommendation could then be provided alongside the CA125 lab result so it is available to the GP when they come to review the report and discuss the result with the patient. Therefore, the information would be provided at a time point when it is of most use to the GP, no information need be self-populated and the risk of prompt fatigue and interference with workflow would be minimal. Qualitative work involving GPs and patients should be undertaken to elucidate their views on the acceptability of options for implementation in order to ensure that the model is implemented in a way which helps rather than hinders clinical decision making. A pilot implementation study could also be performed to ensure that there are no major barriers to its use.[310]

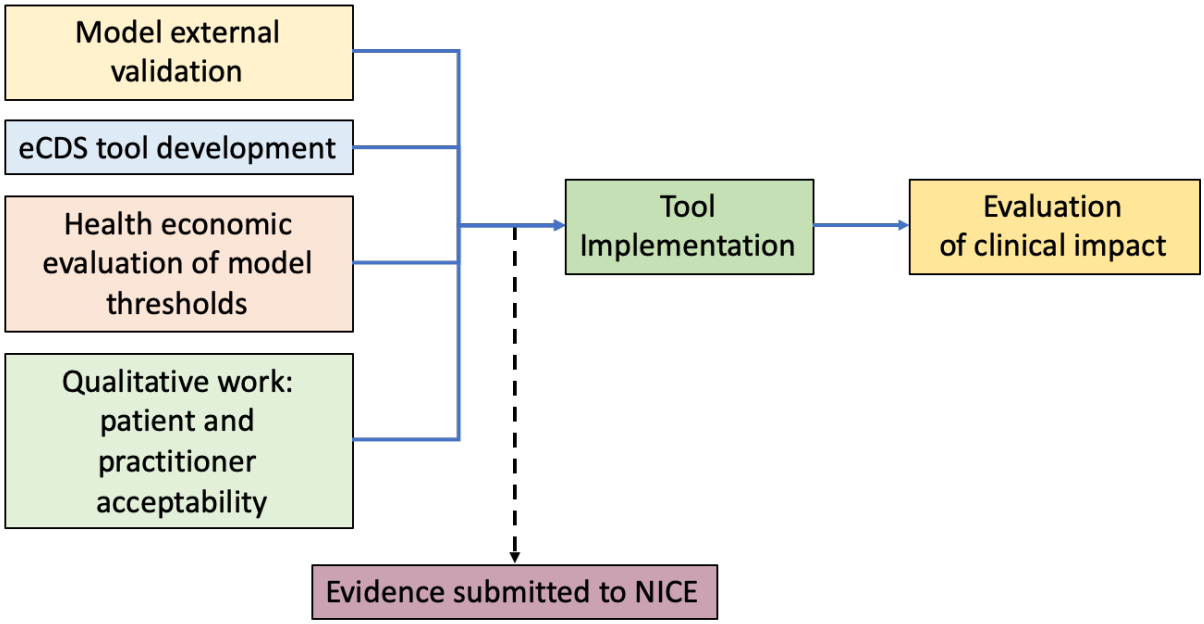


Figure 7.3. Further research to translate the ovarian cancer prediction model into clinical practice.

I plan to develop a grant application to support the further evaluation and implementation of this model as outlined above. If this work is successful, I anticipate submitting evidence to NICE in order for them to consider recommending the model for use as part of the ovarian cancer diagnostic pathway. Following implementation, a mixed methods study could be performed to assess tool uptake and its impact on clinical decision making. Routinely collected primary care and NCRAS datasets could be used to assess the clinical impact of the model on diagnostic interval, stage at diagnosis and survival.[166]

7.4.2 International utility of the model

As demonstrated in the review in **Chapter 1**, CA125 is already used in primary care in a number of countries other than England. Its use is not currently recommended in primary care in some countries due to a lack of evidence on test accuracy, but this thesis provides evidence that CA125 performs well in primary care. The CA125 and age-based ovarian cancer prediction model presented within this thesis could therefore have international utility.

The performance of CA125 and the model may differ in countries where the characteristics of the population are distinct (e.g. different ethnicity) or the test is used differently (e.g. lower threshold for testing) from England. I am currently working with collaborators on a project which will explore the use and accuracy of CA125 in a primary care dataset from the north-west of the United States. If successful, I anticipate collaborating on a study using a large routinely collected data source from the United States, such as that available from the healthcare provider Kaiser Permanente,[311] to validate my diagnostic prediction model.

7.4.3 Post-CA125 testing: non-ovarian cancers

The finding that older women with high CA125 levels in primary care are at high risk of non-ovarian cancer is novel. Therefore, the pathways to diagnosis of women with non-ovarian cancer in this group have not been explored. It is unclear what proportion of women with non-ovarian cancers will be detected on transvaginal ultrasound. It is plausible that having a raised CA125 level could result in diagnostic delay in these women (as they enter the wrong diagnostic pathway) but further research is needed to determine if this is the case. I am currently exploring the potential for using routinely collected primary / secondary care data or an existing dataset from a clinical trial, to examine the sensitivity of ultrasound for non-

ovarian cancers in women with high CA125 levels and to describe their ultimate pathway to diagnosis.

I have presented three potential approaches for the investigation of older women with elevated CA125 levels but without ovarian cancer. Given the high risk of ovarian cancer in these women, I recommend that **Approach a** (recommend GPs consider targeted investigations for other cancers) be implemented in England without delay. **Approach b** (referral to an RDC) could also be adopted as RDCs are rolled out and its impact evaluated within the wider evaluation of RDCs in England.[309] **Approach c** (performing a direct access urgent CT scan) would require further research before implementation to determine whether it is beneficial. Research would need to evaluate the accuracy of CT for non-ovarian cancer in this group. Given the risk of overdiagnosis, the potential for harm (both psychological and physical) from false positive CT scan results and the fact that the majority of women undergoing CT would ultimately not be diagnosed with cancer, it would also be important to ascertain the acceptability and the cost effectiveness of such an approach.

I plan to develop a postdoctoral fellowship application focussed on optimising the post CA125-testing strategy and diagnostic pathway for the detection of both ovarian and non-ovarian cancers.

7.4.4 The best initial test(s)

This thesis has focussed on CA125, as this is the recommended first line test for ovarian cancer in England. However, as highlighted in the review in **Chapter 1**, this approach is not universal. Even within the UK there is variation, with both CA125 and ultrasound recommended in parallel as first line tests in Scotland.[109] There are no published studies which compare different first line tests or test combinations for the detection of ovarian cancer in primary care.[114] Studies conducted in other settings have shown that CA125 is normal in some women and ultrasound abnormal (or *vice versa*) prior to ovarian cancer diagnosis.[55,114,152] Using the tests in parallel is therefore likely to result in better sensitivity but at the cost of specificity, something which was acknowledged by NICE when developing ovarian cancer guidelines in 2011.[114] A comparison of the diagnostic performance of CA125 and ultrasound (alone and in combination) would be useful to help to determine the optimal primary care testing strategy.

CPRD now provide linkage to the HES Diagnostic Imaging Dataset (HES DID) which provides information on what imaging tests have been performed and when.[312] However, it does not include the results of those imaging tests. I am not aware of any available large UK dataset which contains coded or free text information on primary care requested ultrasound results. This makes the evaluation of ultrasound in primary care much more challenging than the evaluation of CA125. A further challenge in using routinely collected data to compare CA125 and ultrasound in England is that ultrasound is the second line test: not all women who have a CA125 test will have had an ultrasound scan. Ideally (if routinely collected data is to be used), the comparison should be made using data from a country, such as Scotland, where both tests are recommended as first line investigations. I am currently exploring available regional datasets within the UK, and also internationally, which could be used to compare the performance of these tests alone and in combination. This may necessitate the use of natural language processing (NLP) techniques to extract outcome information from ultrasound scan reports.[313]

7.4.5 Biomarkers for other cancers

I believe that the work presented within this thesis demonstrates that to understand how a test performs within primary care it must be studied within a primary care population. It is important that this is recognised when translating new biomarkers from higher risk settings to primary care. The recently published CanTest framework provides a 'best practice' guide for researchers, outlining the steps that should be undertaken, from the initial evaluation of a newly discovered test in a high risk population to its implementation and assessment in a low risk population.[314] This could help improve the implementation of cancer diagnostic tests within primary care in the future.

However, CA125 is not the only cancer biomarker to be implemented in primary care without prior evaluation of its diagnostic performance within this setting. Just as CA125 is recommended by NICE for women presenting with symptoms of possible ovarian cancer, PSA is recommended for men presenting with symptoms of possible prostate cancer.[57] In their comprehensive 2015 evidence review, NICE identified a single, small (N=582) study on PSA diagnostic performance in symptomatic men, conducted in the United States, with which to inform their recommendations.[57] PSA is a common primary care test: a recent cohort study

(which used routinely collected data from over 2 million men) found that more than a fifth of men in England (aged 40-74 years) had at least one PSA test performed in primary care with 10 years follow-up.[315] This rose to 44% with 19 years follow-up (1998-2017). An understanding of the diagnostic accuracy of PSA in primary care (both in symptomatic men and in men undergoing opportunistic screening) and knowledge of the probability of prostate cancer at specific PSA levels could help guide decisions on the need for specialist referral. It could also inform guidelines both in the UK and internationally. I intend to apply the methods learnt during my doctoral research to evaluate PSA and develop clinically useful prediction models which incorporate the test level. At the time of thesis submission, a grant proposal (on which I am co-principal investigator) is at the second stage of the review process for an NIHR grant. If successful, the work described in **Chapter 3** and **Chapter 6** will effectively be repeated for PSA. This grant proposal was informed by my experiences during my doctoral research and I have adjusted the methods accordingly. For example, rather than exclude men without a symptom code (as I did during model development in **Chapter 6**) we plan to include all men. This means that we will consider the predictive power of the presence of a symptom *code* within our models, which is appropriate given that eCDS tools are usually auto-populated with coded information drawn from the patient clinical record.

7.5 Conclusion

The research presented within this thesis improves our understanding of the diagnostic performance of CA125 when used in primary care. It demonstrates that, at the conventional cut-off (≥35 U/ml), the test has a high PPV, NPV, specificity and sensitivity for ovarian cancer. However, it also shows that the probability of ovarian cancer varies dramatically depending on a woman's specific CA125 level and age. The probability of ovarian cancer, derived from a relatively simple CA125 and age-based model, could be used in clinical practice to help patients and GPs interpret individual test results and make informed decisions about the need for further investigation. This information could also be used by policy makers to support a risk-based triage system, so that national referral thresholds for ovarian cancer are based on the probability of disease in an individual, rather than using a generic CA125 cut-off.

The novel finding that older women with elevated CA125 levels who do not have ovarian cancer are at high risk of other forms of cancer should change clinical practice both in England

and internationally. While further research is needed to determine the most appropriate approach to investigate these women, it is important that clinicians are aware that elevated CA125 in older women is a multi-cancer risk marker.

References

1 Sung H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; published online Feb 4.

2 Cancer Research UK. Ovarian cancer incidence statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Zero (accessed March 5, 2021).

3 Arnold M, Rutherford MJ, Bardot A, *et al.* Progress in cancer survival, mortality, and incidence in seven high-income countries 1995-2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol* 2019; **20**: 1493–505.

4 Torre LA, Trabert B, DeSantis CE, *et al.* Ovarian cancer statistics, 2018. *CA Cancer J Clin* 2018; **68**: 284–96.

5 Cancer Research UK. Ovarian cancer incidence by age. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One (accessed March 5, 2021).

6 US National Cancer Institute: Surveillance Epidemiology and End Results (SEER) Program. Ovary Cancer: Recent Trends in SEER Age-Adjusted Incidence Rates, 2000-2017. https://seer.cancer.gov/explorer/application.html?site=61&data_type=1&graph_type=2&compareBy=race&chk_race_5=5&chk_race_3=3&chk_race_6=6&chk_race_2=2&hdn_sex=3&age_range=1&stage=101&rate_type=2&advopt_precision=1&advopt_display=2 (accessed March 5, 2021).

7 National Cancer Registration and Analysis Service. Cancer incidence and survival by major ethnic group. 2009. http://www.ncin.org.uk/search/ethnic+group (accessed April 20, 2021).

8 Cancer Research UK. Ovarian cancer incidence by stage at diagnosis. 2017. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Three (accessed March 14, 2021).

9 Cabasag CJ, Butler J, Arnold M, *et al.* Exploring variations in ovarian cancer survival by age and stage (ICBP SurvMark-2): A population-based study. *Gynecol Oncol* 2020; **157**: 234–44.

10 Cancer Research UK. Uterine cancer incidence by stage at diagnosis. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer/incidence#heading-Three (accessed March 14, 2021).

11 Cancer Research UK. Cervical cancer incidence statistics by stage at diagnosis. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/incidence (accessed March 13, 2021).

12 Cancer Research UK. Ovarian cancer mortality statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/mortality (accessed March 15, 2021).

13 Cancer Research UK. Ovarian cancer survival statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three (accessed April 22, 2021).

14 Forbes LJL, Simon AE, Warburton F, *et al.* Differences in cancer awareness and beliefs between Australia, Canada, Denmark, Norway, Sweden and the UK (the International

Cancer Benchmarking Partnership): do they contribute to differences in cancer survival? *Br J Cancer* 2013; **108**: 292–300.
15 Rose PW, Rubin G, Perera-Salazar R, *et al.* Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. *BMJ Open* 2015; **5**: e007212.
16 Cancer Research UK. Ovarian cancer survival by stage at diagnosis. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three (accessed March 14, 2021).
17 Lynch HT, Snyder C, Casey MJ. Hereditary ovarian and breast cancer: what have we learned? *Ann Oncol Off J Eur Soc Med Oncol* 2013; **Suppl 8**: viii83–95.
18 Malander S, Ridderheim M, Måsbäck A, *et al.* One in 10 ovarian cancer patients carry germ line BRCA1 or BRCA2 mutations: results of a prospective study in Southern Sweden. *Eur J Cancer* 2004; **40**: 422–8.
19 Menkiszak J, Gronwald J, Górski B, *et al.* Hereditary ovarian cancer in Poland. *Int J cancer* 2003; **106**: 942–5.
20 Antoniou A, Pharoah PDP, Narod S, *et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003; **72**: 1117–30.
21 Kuchenbaecker KB, Hopper JL, Barnes DR, *et al.* Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA - J Am Med Assoc* 2017; **317**: 2402–16.
22 Cancer Research UK. Lifetime risk of cancer. https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk#heading-One (accessed April 4, 2021).
23 Ryan NA, McMahon RF, Ramchander NC, Seif MW, Evans DG, Crosbie EJ. Lynch syndrome for the gynaecologist. *Obstet Gynaecol* 2021; **23**: 9–20.
24 Watson P, Vasen HFA, Mecklin JP, *et al.* The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int J Cancer* 2008; **123**: 444–9.
25 Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. *Cancer Biol Med* 2017; **14**: 9–32.
26 Kazerouni N, Greene MH, Lacey J V., Mink PJ, Schairer C. Family history of breast cancer as a risk factor for ovarian cancer in a prospective study. *Cancer* 2006; **107**: 1075–83.
27 Hemminki K, Sundquist J, Brandt A. Incidence and mortality in epithelial ovarian cancer by family history of any cancer. *Cancer* 2011; **117**: 3972–80.
28 Gong TT, Wu QJ, Vogtmann E, Lin B, Wang YL. Age at menarche and risk of ovarian cancer: A meta-analysis of epidemiological studies. *Int J Cancer* 2013; **132**: 2894–2900.
29 Tsilidis KK, Allen NE, Key TJ, *et al.* Oral contraceptive use and reproductive factors and risk of ovarian cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer* 2011; **105**: 1436–42.
30 Beral V, Doll R, Hermon C, *et al.* Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23 257 women with ovarian cancer and 87 303 controls. *Lancet* 2008; **371**: 303–14.
31 Sung HK, Ma SH, Choi J-Y, *et al.* The Effect of Breastfeeding Duration and Parity on the Risk of Epithelial Ovarian Cancer: A Systematic Review and Meta-analysis. *J Prev Med Public Health* 2016; **49**: 349–66.

32 Gaitskell K, Green J, Pirie K, *et al.* Histological subtypes of ovarian cancer associated with parity and breastfeeding in the prospective Million Women Study. *Int J Cancer* 2018; **142**: 281–9.
33 Soegaard M, Kjaer SK, Cox M, *et al.* Menopausal hormone use and ovarian cancer risk: Individual participant meta-analysis of 52 epidemiological studies. *Lancet* 2015; **385**: 9–15.
34 Liu Z, Zhang TT, Zhao JJ, *et al.* The association between overweight, obesity and ovarian cancer: A meta-analysis. *Jpn J Clin Oncol* 2015; **45**: 1107–15.
35 Bhaskaran K, Douglas I, Forbes H, Dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: A population-based cohort study of 5·24 million UK adults. *Lancet* 2014; **384**: 755–65.
36 Aune D, Rosenblatt DAN, Chan DSM, *et al.* Anthropometric factors and ovarian cancer risk: A systematic review and nonlinear dose-response meta-analysis of prospective studies. *Int J Cancer* 2015; **136**: 1888–98.
37 Beral V, Hermon C, Peto R, *et al.* Ovarian cancer and body size: Individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS Med* 2012; **9**: e1001200.
38 Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol* 2011; **12**: 785–94.
39 Constanza Camargo M, Stayner LT, Straif K, *et al.* Occupational exposure to asbestos and ovarian cancer: A meta-analysis. *Environ Health Perspect* 2011; **119**: 1211–7.
40 Faber MT, Kjær SK, Dehlendorff C, *et al.* Cigarette smoking and risk of ovarian cancer: A pooled analysis of 21 case-control studies. *Cancer Causes Control* 2013; **24**: 989–1004.
41 International Agency for Research on Cancer. List of classifications by cancer sites with sufficient or limited evidence in humans, IARC Monographs Volumes 1–129a. 2020.
42 O'Brien KM, Tworoger SS, Harris HR, *et al.* Association of Powder Use in the Genital Area With Risk of Ovarian Cancer. *JAMA* 2020; **323**: 49–59.
43 Matz M, Coleman MP, Sant M, *et al.* The histology of ovarian cancer: worldwide distribution and implications for international survival comparisons (CONCORD-2). *Gynecol Oncol* 2017; **144**: 405–13.
44 Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, Karlan BY. Ovarian cancer. *Nat Rev Dis Prim* 2016; **2**: 16061.
45 Köbel M, Kalloger SE, Huntsman DG, *et al.* Differences in tumor type in low-stage versus high-stage ovarian carcinomas. *Int J Gynecol Pathol* 2010; **29**: 203–11.
46 National cancer intelligence netwrok (NCIN). A profile of ovarian cancerin England. London, 2012.
47 Royal College of Obstetricians and Gynaecologists. Management of Female Malignant Ovarian Germ Cell Tumours. London, 2016.
48 Taylor HC. Malignant and semi-malignant tumors of the ovary. *Surg Gynecol Obs* 1929; **48**: 204–30.
49 Du Bois A, Ewald-Riegler N, De Gregorio N, *et al.* Borderline tumours of the ovary: A cohort study of the Arbeitsgmeinschaft Gynäkologische Onkologie (AGO) Study Group. *Eur J Cancer* 2013; **49**: 1905–14.
50 Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO Classification of Tumours

of Female Reproductive Organs. Fourth Edition. Lyon: International Agency for Research on Cancer, 2014.
51 World Health Organisation. International Classification of Diseases for Oncology (ICD-O). Third edition. Geneva, 2013.
52 FIGO Committee on Gynecologic Oncology. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *Int J Gynaecol Obstet* 2014; **124**: 1–5.
53 Pinsky PF, Yu K, Kramer BS, *et al.* Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. *Gynecol Oncol* 2016; **143**: 270–5.
54 Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *BMJ* 2012; **344**: d8009–d8009.
55 Gilbert L, Basso O, Sampalis J, *et al.* Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *Lancet Oncol* 2012; **13**: 285–91.
56 Gershenson DM. Management of borderline ovarian tumours. *Best Pract Res Clin Obstet Gynaecol* 2017; **41**: 49–59.
57 National Institute for Health and Care Excellence. Suspected cancer: recognition and referral (NG12). 2015.
58 Dubeau L. The cell of origin of ovarian epithelial tumors and the ovarian surface epithelium dogma: Does the emperor have no clothes? Gynecol. Oncol. 1999; **72**: 437–42.
59 Reade CJ, McVey RM, Tone AA, *et al.* The Fallopian Tube as the Origin of High Grade Serous Ovarian Cancer: Review of a Paradigm Shift. *J Obstet Gynaecol Canada* 2014; **36**: 133–40.
60 Kindelberger DW, Lee Y, Miron A, *et al.* Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. *Am J Surg Pathol* 2007; **31**: 161–9.
61 Przybycin CG, Kurman RJ, Ronnett BM, Shih IM, Vang R. Are all pelvic (Nonuterine) serous carcinomas of tubal origin? *Am J Surg Pathol* 2010; **34**: 1407–16.
62 Labidi-Galy SI, Papp E, Hallberg D, *et al.* High grade serous ovarian carcinomas originate in the fallopian tube. *Nat Commun* 2017; **8**.
63 Zhang S, Dolgalev I, Zhang T, Ran H, Levine DA, Neel BG. Both fallopian tube and ovarian surface epithelium are cells-of-origin for high-grade serous ovarian carcinoma. *Nat Commun* 2019; **10**: 5367.
64 National Cancer Institute (NCI). Ovarian Epithelial, Fallopian Tube, and Primary Peritoneal Cancer Treatment (PDQ®)–Health Professional Version. https://www.cancer.gov/types/ovarian/hp/ovarian-epithelial-treatment-pdq#cit/section_1.3 (accessed March 8, 2021).
65 National Institute for Health and Clinical Excellence. Ovarian cancer: recognition and initial management (CG122). 2011. https://www.nice.org.uk/guidance/cg122/resources/ovarian-cancer-recognition-and-initial-management-35109446543557 (accessed April 17, 2021).
66 Hiom SC. Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *Br J Cancer* 2015; **112**: S1–5.
67 Crosby D, Lyons N, Greenwood E, *et al.* A roadmap for the early detection and diagnosis of cancer. *Lancet Oncol* 2020; **21**: 1397–9.
68 National Health Service. The NHS Long term plan. London, 2019.

69 Baldwin LA, Huang B, Miller RW, *et al.* Ten-Year Relative Survival for Epithelial Ovarian Cancer. *Obstet Gynecol* 2012; **120**: 612–8.
70 Vergote I, Tropé CG, Amant F, *et al.* Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *N Engl J Med* 2010; **363**: 943–53.
71 Chi DS, Eisenhauer EL, Lang J, *et al.* What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? *Gynecol Oncol* 2006; **103**: 559–64.
72 Chang S-J, Hodeib M, Chang J, Bristow RE. Survival impact of complete cytoreduction to no gross residual disease for advanced-stage ovarian cancer: A meta-analysis. *Gynecol Oncol* 2013; **130**: 493–8.
73 National Cancer Registration & Analysis Service. Get data out: Ovary, fallopian tube and primary peritoneal carcinomas. 2020. https://www.cancerdata.nhs.uk/getdataout/ovary (accessed March 12, 2021).
74 Ovarian Cancer Audit Feasibility Pilot Steering Group. Disease Profile in England: Incidence, mortality, stage and survival for ovary, fallopian tube and primary peritoneal carcinomas. London, 2020.
75 Love AJ, Lambert P, Turner D, *et al.* Diagnostic and referral intervals for Manitoba women with epithelial ovarian cancer - the Manitoba Ovarian Cancer Outcomes (MOCO) study group: a retrospective cross-sectional study. *CMAJ Open* 2017; **5**: E116–22.
76 Tørring ML, Frydenberg M, Hansen RP, Olesen F, Vedsted P. Evidence of increasing mortality with longer diagnostic intervals for five common cancers: A cohort study in primary care. *Eur J Cancer* 2013; **49**: 2187–98.
77 Tørring ML, Falborg AZ, Jensen H, *et al.* Advanced-stage cancer and time to diagnosis: An International Cancer Benchmarking Partnership (ICBP) cross-sectional study. *Eur J Cancer Care (Engl)* 2019; **28**: e13100.
78 Nagle CM, Francis JE, Nelson AE, *et al.* Reducing Time to Diagnosis Does Not Improve Outcomes for Women With Symptomatic Ovarian Cancer: A Report From the Australian Ovarian Cancer Study Group. *J Clin Oncol* 2011; **29**: 2253–8.
79 Neal RD, Tharmanathan P, France B, *et al.* Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 2015; **112**: S92–107.
80 Round T, Gildea C, Ashworth M, Møller H. Association between use of urgent suspected cancer referral and mortality and stage at diagnosis: a 5-year national cohort study. *Br J Gen Pract* 2020; **70**: e389 LP-e398.
81 Robinson KM, Christensen KB, Ottesen B, Krasnik A. Diagnostic delay, quality of life and patient satisfaction among women diagnosed with endometrial or ovarian cancer: a nationwide Danish study. *Qual Life Res* 2012; **21**: 1519–25.
82 Price S, Spencer A, Zhang X, *et al.* Trends in time to cancer diagnosis around the period of changing national guidance on referral of symptomatic patients: A serial cross-sectional study using UK electronic healthcare records from 2006–17. *Cancer Epidemiol* 2020; **69**: 101805.
83 Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: Findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012; **13**: 353–65.
84 Swann R, McPhail S, Shand B, *et al.* Diagnosing cancer in primary care: Results from

the National Cancer Diagnosis Audit. *Br J Gen Pract* 2018; **68**: e63–72.
85 Buys SS, Partridge E, Black A, *et al.* Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* 2011; **305**: 2295–303.
86 Kobayashi H, Yamada Y, Sado T, *et al.* A randomized study of screening for ovarian cancer: a multicenter study in Japan. *Int J Gynecol Cancer* 2008; **18**: 414–20.
87 Barrett J, Sharp DJ, Stapley S, Stabb C, Hamilton W. Pathways to the diagnosis of ovarian cancer in the UK: A cohort study in primary care. *BJOG* 2010; **117**: 610–4.
88 National Cancer Registration and Analysis Service. Routes to diagnosis: data overview (2006-2017). London, 2020.
89 Ebell MH, Culp MB, Radke TJ. A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer. *Am J Prev Med* 2016; **50**: 384–94.
90 Chalmers V, Allen V. Ovarian cancer hope as experts move one step closer to a screening method after discovery of six types of cell that trigger the 'silent killer'. Dly. Mail. 2020; published online Feb 10. https://www.dailymail.co.uk/health/article-7987639/Ovarian-cancer-hope-discovery-six-types-cell-trigger-silent-killer.html.
91 Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* 2009; **339**: b2998.
92 Goff BA, Mandel LS, Melancon CH, Muntz HG. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA* 2004; **291**: 2705–12.
93 Lurie G, Wilkens LR, Thompson PJ, Matsuno RK, Carney ME, Goodman MT. Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: A case analysis. *Gynecol Oncol* 2010; **119**: 278–84.
94 Koo MM, Swann R, McPhail S, *et al.* Presenting symptoms of cancer and stage at diagnosis: evidence from a cross-sectional, population-based study. *Lancet Oncol* 2020; **21**: 73–9.
95 Lim AWW, Mesher D, Sasieni P. Estimating the workload associated with symptoms-based ovarian cancer screening in primary care: An audit of electronic medical records. *BMC Fam Pract* 2014; **15**: 200.
96 Dutch Society For Obstetrics and Gynaecology. Epithelial ovarian carcinoma. 2018. www.oncoline.nl/ovariumcarcinoom (accessed July 12, 2018).
97 Cancer Australia. Assessment of symptoms that may be ovarian cancer: a guide for general practitioners. 2015. https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/assessment-symptoms-may-be-ovarian-cancer-guide-gps (accessed Sept 12, 2019).
98 Cancer Australia. Appropriate referral of women with suspected ovarian cancer. 2015. https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/appropriate-referral-women-suspected-ovarian-cancer (accessed July 12, 2018).
99 Cancer Council Victoria. Optimal care pathway for women with ovarian cancer. 2015. https://www.cancervic.org.au/downloads/health-professionals/optimal-care-pathways/Optimal_care_pathway_for_women_with_ovarian_cancer.pdf (accessed Aug 31, 2018).
100 New Zealand Guidelines Group. Suspected Cancer in Primary Care: Guidelines for investigation, referral and reducing ethnic disparity. 2009. https://www.health.govt.nz/publication/suspected-cancer-primary-care-guidelines-

investigation-referral-and-reducing-ethnic-disparities (accessed July 12, 2018).
101 Society of Gynecologic Oncology. Committee Opinion No. 716: The Role of the Obstetrician-Gynecologist in the Early Detection of Epithelial Ovarian Cancer in Women at Average Risk. *Obstet Gynecol* 2017; **130**: e146–9.
102 Cancer Care Ontario. Ovarian cancer diagnosis pathway map. 2016. www.cancercareontario.ca/en/pathway-maps/ovarian-cancer (accessed July 12, 2018).
103 National Comprehensive Caner Network. Ovarian cancer: including fallopian tube cancer and primary peritoneal cancer. 2018. https://www.nccn.org/professionals/physician_gls/default.aspx#ovarian (accessed Aug 31, 2018).
104 Guidelines and Protocol Advisory Comittee, British Columbia Medical Services Commission. Genital tract cancers in females: ovarian, fallopian tube, and primary peritoneal cancers. 2014. https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/bc-guidelines/ovarian-cancers (accessed July 12, 2018).
105 German Guideline Programme in Oncology (OL) of the Association of the Scientific Medical Societies in Germany (AWMF). Guideline on diagnostics, therapy and follow-up of malignant ovarian tumours. 2017. https://www.awmf.org/leitlinien/detail/ll/032-035OL.html (accessed Aug 4, 2018).
106 British Gynaecological Cancer Society. Epithelial ovarian / fallopian tube / primary peritoneal cancer guidelines: recommendations for practice. 2017. bgcs.org.uk/BGCS Guidelines Ovarian Guidelines 2017.pdf (accessed July 12, 2018).
107 Ireland National Cancer Control Programme. Ovarian cancer GP referral for symptomatic women. 2016. www.hse.ie/eng/services/list/5/cancer/profinfo/resources/gpreferrals/Ovarian-Cancer-Referral-Guidelines.pdf (accessed Sept 12, 2019).
108 Healthcare Improvement Scotland. Scottish Referral Guidelines for Suspected Cancer. 2014. http://www.healthcareimprovementscotland.org/our_work/cancer_care_improvement/programme_resources/scottish_referral_guidelines.aspx (accessed July 12, 2018).
109 Scottish Intercollegiate Guidelines Network. Management of epithelial ovarian cancer. 2013. http://www.sign.ac.uk/sign-135-management-of-epithelial-ovarian-cancer.html (accessed July 12, 2018).
110 The Danish National Health Authority. Integrated ovarian cancer patient pathway. 2016. https://www.sst.dk/da/udgivelser/2016/~/media/765a142bc73446399c769a5d8b54c3dc.ash x (accessed July 12, 2018).
111 The Norwegian Directorate of Health. Ovarian cancer patient pathway. 2016. https://helsedirektoratet.no/Retningslinjer/Pakkeforløp for eggstokkreft.pdf (accessed Aug 29, 2018).
112 Regional Cancer Centre Co-operative Sweden. Standardised Ovarian Cancer Care Pathway. 2015. https://www.cancercentrum.se/globalassets/cancerdiagnoser/gynekologi/aggstockscancer/vardforlopp/standardiserat_vardforlopp_aggstockar_20151221.pdf (accessed Aug 25, 2018).
113 mapchart.net. World custom maps. https://mapchart.net (accessed March 16, 2021).
114 National Collaborating Centre for Cancer. Ovarian Cancer: The recognition and initial

management of ovarian cancer (NICE Clinical Guidelines No. 122). Cardiff (UK), 2011.
115 Funston G, O'Flynn H, Ryan NAJ, Hamilton W, Crosbie EJ. Recognizing Gynecological Cancer in Primary Care: Risk Factors, Red Flags, and Referrals. *Adv Ther* 2018; **35**: 577–589.
116 Funston G, Melle VM, Ladegaard Baun M-L, *et al.* Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. *BMC Cancer* 2019; **19**: 1028.
117 Bast RC, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest* 1981; **68**: 1331–1337.
118 Bast RC, Klug TL, John E St., *et al.* A Radioimmunoassay Using a Monoclonal Antibody to Monitor the Course of Epithelial Ovarian Cancer. *N Engl J Med* 1983; **309**: 883–7.
119 Jacobs I, Bast RC. The CA 125 tumour-associated antigen: A review of the literature. Hum. Reprod. 1989; **4**: 1–12.
120 Duffy MJ, Bonfrer JM, Kulpa J, *et al.* CA125 in ovarian cancer: European Group on Tumor Markers guidelines for clinical use. *Int J Gynecol Cancer* 2005; **15**: 679–91.
121 Haridas D, Ponnusamy MP, Chugh S, Lakshmanan I, Seshacharyulu P, Batra SK. MUC16: molecular analysis and its functional implications in benign and malignant conditions. *FASEB J* 2014; **28**: 4183–99.
122 Felder M, Kapur A, Gonzalez-Bosquet J, *et al.* MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Mol Cancer* 2014; **13**: 129.
123 Kolwijck E, Thomas CMG, Bulten J, Massuger LFAG. Preoperative CA-125 levels in 123 patients b with borderline ovarian tumors a retrospective analysis and review of the literature. *Int J Gynecol Cancer* 2009; **19**: 1335–8.
124 Kawai M, Kano T, Kikkawa F, *et al.* Seven tumor markers in benign and malignant germ cell tumors of the ovary. *Gynecol Oncol* 1992; **45**: 248–53.
125 Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. Accuracy of CA 125 in the diagnosis of ovarian tumors: A quantitative systematic review. *Eur J Obstet Gynecol Reprod Biol* 2009; **142**: 99–105.
126 Duraker N, Hot S, Polat Y, Höbek A, Gençler N, Urhan N. CEA, CA 19-9, and CA 125 in the differential diagnosis of benign and malignant pancreatic diseases with or without jaundice. *J Surg Oncol* 2007; **95**: 142–7.
127 Cwik G, Wallner G, Skoczylas T, Ciechanski A, Zinkiewicz K. Cancer antigens 19-9 and 125 in the differential diagnosis of pancreatic mass lesions. *Arch Surg* 2006; **141**: 968–73.
128 Haglund C. Tumour marker antigen CA125 in pancreatic cancer: A comparison with CA19-9 and CEA. *Br J Cancer* 1986; **54**: 897–901.
129 Molina R, Auge JM, Escudero JM, *et al.* Mucins CA 125, CA 19.9, CA 15.3 and TAG-72.3 as tumor markers in patients with lung cancer: comparison with CYFRA 21-1, CEA, SCC and NSE. *Tumour Biol* 2008; **29**: 371–80.
130 Ginath S, Menczer J, Fintsi Y, Ben-Shem E, Glezerman M, Avinoach I. Tissue and serum CA125 expression in endometrial cancer. *Int J Gynecol Cancer* 2002; **12**: 372–5.
131 Chao A, Tang Y-H, Lai C-H, *et al.* Potential of an age-stratified CA125 cut-off value to improve the prognostic classification of patients with endometrial cancer. *Gynecol Oncol* 2013; **129**: 500–4.
132 Luo H, Shen K, Li B, Li R, Wang Z, Xie Z. Clinical significance and diagnostic value of serum NSE, CEA, CA19-9, CA125 and CA242 levels in colorectal cancer. *Oncol Lett*

2020; **20**: 742–50.
133 Zhong W, Yu Z, Zhan J, *et al.* Association of serum levels of CEA, CA199, CA125, CYFRA21-1 and CA72-4 and disease characteristics in colorectal cancer. *Pathol Oncol Res* 2015; **21**: 83–95.
134 Gao Y, Wang J, Zhou Y, Sheng S, Qian SY, Huo X. Evaluation of Serum CEA, CA19-9, CA72-4, CA125 and Ferritin as Diagnostic Markers and Factors of Clinical Parameters for Colorectal Cancer. *Sci Rep* 2018; **8**: 2732.
135 Mol BWJ, Bayram N, Lijmer JG, *et al.* The performance of CA-125 measurement in the detection of endometriosis: A meta-analysis. *Fertil Steril* 1998; **70**: 1101–8.
136 Moore RG, Miller MC, Steinhoff MM, *et al.* Serum HE4 levels are less frequently elevated than CA125 in women with benign gynecologic disorders. *Am J Obstet Gynecol* 2012; **206**: 351.e1-351.e8.
137 Halila H, Stenman U-H, Seppälä M. Ovarian cancer antigen CA 125 levels in pelvic inflammatory disease and pregnancy. *Cancer* 1986; **57**: 1327–9.
138 Malkasian GD, Knapp RC, Lavin PT, *et al.* Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. *Am J Obstet Gynecol* 1988; **159**: 341–6.
139 Yilmaz MB, Nikolaou M, Cohen Solal A. Tumour biomarkers in heart failure: is there a role for CA-125? *Eur J Heart Fail* 2011; **13**: 579–83.
140 Faggiano P, D'Aloia A, Brentana L, *et al.* Serum levels of different tumour markers in patients with chronic heart failure. *Eur J Heart Fail* 2005; **7**: 57–61.
141 Devarbhavi H, Kaese D, Williams AW, Rakela J, Klee GG, Kamath PS. Cancer Antigen 125 in Patients With Chronic Liver Disease. *Mayo Clin Proc* 2002; **77**: 538–41.
142 Fang C, Cao Y, Liu X, Zeng X-T, Li Y. Serum CA125 is a predictive marker for breast cancer outcomes and correlates with molecular subtypes. *Oncotarget* 2017; **8**: 63963–70.
143 Bon GG, Kenemans P, Verstraeten R, Van Kamp GJ, Hilgers J. Serum tumor marker immunoassays in gynecologic oncology: establishment of reference values. *Am J Obstet Gynecol* 1996; **171**: 107–14.
144 Han SN, Lotgerink A, Gziri MM, Van Calsteren K, Hanssens M, Amant F. Physiologic variations of serum tumor markers in gynecological malignancies during pregnancy: a systematic review. *BMC Med* 2012; **10**: 86.
145 Grover S, Koh H, Weideman P, Quinn MA. The effect of the menstrual cycle on serum CA 125 levels: A population study. *Am J Obstet Gynecol* 1992; **167**: 1379–1381.
146 Chen F-P, Soong Y-K, Lee N, Kai Lo S. The use of serum CA-125 as a marker for endometriosis in patients with dysmenorrhea for monitoring therapy and for recurrence of endometriosis. *Acta Obstet Gynecol Scand* 1998; **77**: 665.
147 Monteiro S, Franco F, Costa S, *et al.* Prognostic value of CA125 in advanced heart failure patients. *Int J Cardiol* 2010; **140**: 115–8.
148 Meng Q, Shi S, Liang C, *et al.* Diagnostic accuracy of a CA125-based biomarker panel in patients with pancreatic cancer: A systematic review and meta-analysis. *J Cancer* 2017; **8**: 3615–22.
149 Cohen JD, Li L, Wang Y, *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018; **359**: 926 LP – 930.
150 Graham L. ACOG releases guidelines on management of adnexal masses. *Am Fam Physician* 2008; **77**: 1320–3.

151 Hartman CA, Juliato CRT, Sarian LO, *et al.* Multimodal approach to screening for ovarian cancer. *Ultrasound Obstet Gynecol* 2012; **40**: 360–6.
152 Buys SS, Partridge E, Greene MH, *et al.* Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: Findings from the initial screen of a randomized trial. *Am J Obstet Gynecol* 2005; **193**: 1630–9.
153 Menon U, Gentry-Maharaj A, Hallett R, *et al.* Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 2009; **10**: 327–40.
154 Jacobs IJ, Menon U, Ryan A, *et al.* Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* 2016; **387**: 945–56.
155 National Institute for Health and Care Excellence. Surveillance report on 'suspected cancer: recognition and referral (NICE guideline NG12)'. London, 2020 https://www.nice.org.uk/guidance/ng12/resources/2020-surveillance-of-suspected-cancer-recognition-and-referral-nice-guideline-ng12-pdf-9170799052741.
156 Crawford SM, Evans C. Outcome of elevated CA125 values from primary care following implementation of ovarian cancer guidelines. *Fam Pract* 2017; **28**: 28.
157 Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* 2016; **353**: i3139.
158 Lutgendorf MA, Stoll KA. Why 99% may not be as good as you think it is: limitations of screening for rare diseases. *J Matern Neonatal Med* 2016; **29**: 1187–9.
159 Nicholson BD, Lee M-M, Wijeratne D, James T, Shine B, Oke JL. Trends in Cancer Antigen 125 testing 2003-2014: A primary care population-based cohort study using laboratory data. *Eur J Cancer Care (Engl)* 2018; **0**: e12914.
160 Funston G, Zhou Y, Price S, *et al.* The association between CA125 level, diagnostic interval and stage at diagnosis in ovarian cancer: an analysis of CPRD and NCRAS data. In: The Cancer and Primary Care International Network (Ca-PRI) Conference. Toronto, 2019.
161 Nicholson B, Mant D, Shinkins B, *et al.* International variation in adherence to referral guidelines for suspected cancer: A secondary analysis of survey data. *Br J Gen Pract* 2016; **66**: e106–13.
162 Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med* 2015; **162**: W1-73.
163 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ* 2017; **357**: j2099.
164 National Institute for Health and Care Excellence. CVD risk assessment and management (Clinical Knowledge Summary). 2020. https://cks.nice.org.uk/topics/cvd-risk-assessment-management/ (accessed Feb 28, 2021).
165 Hippisley-Cox J, Copeland C. QCancer Tool. https://qcancer.org (accessed Aug 9, 2020).
166 Hamilton W. Electronic Risk Assessment for Cancer for Patients in General Practice (ISRCTN22560297). ISRCTN Regist. http://www.isrctn.com/ISRCTN22560297 (accessed Oct 30, 2020).

167 Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. *Br J Cancer* 2015; **113**: 1645–50.
168 Moore RG, McMeekin DS, Brown AK, *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009; **112**: 40–6.
169 Louie KS, Seigneurin A, Cathcart P, Sasieni P. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. *Ann Oncol* 2015; **26**: 848–64.
170 Bossuyt PM, Reitsma JB, Bruns DE, *et al.* STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351**: h5527.
171 Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating, 2nd edn. Springer New York LLC, 2019.
172 Andersen MR, Lowe KA, Goff BA. Value of Symptom-Triggered Diagnostic Evaluation for Ovarian Cancer. *Obstet Gynecol* 2014; **123**: 73–9.
173 Henson KE, Elliss-Brookes L, Coupland VH, *et al.* Data Resource Profile: National Cancer Registration Dataset in England. *Int J Epidemiol* 2020; **49**: 16-16h.
174 Bailey SER, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: A prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract* 2017; **67**: e405–13.
175 Watson J, Salisbury C, Banks J, Whiting P, Hamilton W. Predictive value of inflammatory markers for cancer diagnosis in primary care: a prospective cohort study using electronic health records. *Br J Cancer* 2019; **120**: 1045–51.
176 Nicholson BD, Aveyard P, Price SJ, Hobbs FR, Koshiaris C, Hamilton W. Prioritising primary care patients with unexpected weight loss for cancer investigation: diagnostic accuracy study. *BMJ* 2020; **370**: m2651–m2651.
177 Bradley SH, Hatton NLF, Aslam R, *et al.* Estimating lung cancer risk from chest X-ray and symptoms: a prospective cohort study. *Br J Gen Pract* 2021; **71**: e280 LP-e286.
178 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**: 827–36.
179 Wolf A, Dedman D, Campbell J, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019; **48**: 1740–1740g.
180 QResearch. The QResearch database. https://www.qresearch.org (accessed Jan 8, 2019).
181 Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* 2009; **339**: b2998.
182 Lijmer JG, Mol BW, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; **282**: 1061–6.
183 Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; **174**: 469–76.
184 Whiting PF, Rutjes AWS, Westwood ME, *et al.* Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36.
185 The Clinical Practice Research Datalink. CPRD Bibliography. https://www.cprd.com/bibliography (accessed Jan 14, 2021).
186 Chisholm J. The Read clinical classification. *Br Med J* 1990; **300**: 1092.
187 Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text

records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case-control study. *BMJ Open* 2016; **6**: e011664.
188 Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol* 2019; **34**: 91–9.
189 Public Health England. National Cancer Registration and Analysis Service (NCRAS). 2016. https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras (accessed July 18, 2019).
190 Union for International Cancer Control. TNM Classification of Malignant Tumours, 8th edn. Wiley Blackwell, 2017.
191 Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; **46**: 1093–1093i.
192 The Clinical Practice Research Datalink. Hospital Episode Statistics (HES) Admitted Patient Care and CPRD primary care data Documentation (Set 16), V2.1. London, 2018.
193 Townsend P, Phillimore P, Beattie A. Health and deprivation: Inequality and the North. London: Croom Helm, 1987.
194 The Clinical Practice Research Datalink. Small area level data based on patient postcode, Version 2.4. London, 2018.
195 Badrick E, Renehan I, Renehan AG. Linkage of the UK Clinical Practice Research Datalink with the national cancer registry. *Eur J Epidemiol* 2019; **34**: 101–2.
196 Arhi CS, Bottle A, Burns EM, *et al.* Comparison of cancer diagnosis recording between the Clinical Practice Research Datalink, Cancer Registry and Hospital Episodes Statistics. *Cancer Epidemiol* 2018; **57**: 148-157.
197 Boggon R, Van Staa TP, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiol Drug Saf* 2013; **22**: 168–75.
198 Strongman H, Williams R, Bhaskaran K. What are the implications of using individual and combined sources of routinely collected data to identify and characterise incident site-specific cancers? a concordance and validation study using linked English electronic health records data. *BMJ Open* 2020; **10**: e037719.
199 World Health Organisation. International Classification of Diseases for Oncology (ICD-O). Second edition. Geneva: World Health Organisation, 1990.
200 Banks J, Hollinghurst S, Bigwood L, Peters TJ, Walter FM, Hamilton W. Preferences for cancer investigation: A vignette-based study of primary-care attendees. *Lancet Oncol* 2014; **15**: 232–40.
201 Haga Y, Sakamoto K, Egami H, Yoshimura R, Akagi M. Evaluation of serum CA125 values in healthy individuals and pregnant women. *Am J Med Sci* 1986; **292**: 25–9.
202 Pauler DK, Menon U, McIntosh M, Symecko HL, Skates SJ, Jacobs IJ. Factors influencing serum CA125II levels in healthy postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2001; **10**: 489–93.
203 National Institute for Health and Care Excellence. Menopause: diagnosis and management (NG23). 2015.
204 Fischerova D, Zikan M, Dundr P, Cibula D. Diagnosis, treatment, and follow-up of borderline ovarian tumors. *Oncologist* 2012; **17**: 1515–33.

205 Seed P. DIAGT: Stata module to report summary statistics for diagnostic tests compared to true disease status. Stat. Softw. Components S423401, Bost. Coll. Dep. Econ. 2010.
206 Stata. roctab — Nonparametric ROC analysis. https://www.stata.com/manuals/rroctab.pdf (accessed March 1, 2021).
207 Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J Thorac Oncol* 2010; **5**: 1315–6.
208 Harrell FE. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. New York: Springer, 2001.
209 Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. Research Data Supporting 'The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study'. Cambridge Univ. Repos. 2020. DOI:https://doi.org/10.17863/CAM.56363.
210 Johnson CC, Kessel B, Riley TL, *et al.* The epidemiology of CA-125 in women without evidence of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial. *Gynecol Oncol* 2008; **110**: 383–9.
211 Sjövall K, Nilsson B, Einhorn N. The significance of serum CA 125 elevation in malignant and nonmalignant diseases. *Gynecol Oncol* 2002; **85**: 175–8.
212 Olaitan A. Rapid response to: Recognition and initial management of ovarian cancer: summary of NICE guidance. BMJ. 2011. https://www.bmj.com/rapidresponse/%0D2011/11/03/recognition-and-initial-management-ovarian-cancersummary-%0Dnice-guidance (accessed April 9, 2021).
213 Tate AR, Nicholson A, Cassell JA. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010; **102**: 947–51.
214 Barclay ME, Abel GA, Greenberg DC, Rous B, Lyratzopoulos G. Socio-demographic variation in stage at diagnosis of breast, bladder, colon, endometrial, lung, melanoma, prostate, rectal, renal and ovarian cancer in England and its population impact. *Br J Cancer* 2021; **124**: 1320–1329.
215 Goff BA, Mandel LS, Drescher CW, *et al.* Development of an ovarian cancer symptom index: Possibilities for earlier detection. *Cancer* 2007; **109**: 221–7.
216 Bast RC, Lu Z, Han CY, *et al.* Biomarkers and Strategies for Early Detection of Ovarian Cancer. *Cancer Epidemiol Biomarkers & Prev* 2020; **29**: 2504 LP – 2512.
217 Mounce LTA, Price S, Valderas JM, Hamilton W. Comorbid conditions delay diagnosis of colorectal cancer: a cohort study using electronic primary care records. *Br J Cancer* 2017; **116**: 1536–43.
218 Jensen H, Tørring ML, Olesen F, Overgaard J, Fenger-Grøn M, Vedsted P. Diagnostic intervals before and after implementation of cancer patient pathways – a GP survey and registry based comparison of three cohorts of cancer patients. *BMC Cancer* 2015; **15**: 308.
219 Walter FM, Mills K, Mendonça SC, *et al.* Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. *Lancet Gastroenterol Hepatol* 2016; **1**: 298–306.
220 Ridd MJ, Santos Ferreira DL, Montgomery AA, Salisbury C, Hamilton W. Patient-doctor continuity and diagnosis of cancer: Electronic medical records study in general practice. *Br J Gen Pract* 2015; **65**: e305–e311.
221 Stata manual. streg - parametric survival models.

https://www.stata.com/manuals/ststreg.pdf (accessed Feb 2, 2021).
222 Stata manual. Wald test of linear hypotheses. https://www.stata.com/manuals/semtest.pdf (accessed Feb 2, 2021).
223 McDonald J. Handbook of Biological Statistics (3rd Edition). Baltimore: Sparky House Publishing, 2014.
224 Lyratzopoulos G, Abel GA, Brown CH, *et al.* Socio-demographic inequalities in stage of cancer diagnosis: evidence from patients with female breast, lung, colon, rectal, prostate, renal, bladder, melanoma, ovarian and endometrial cancer. *Ann Oncol* 2013; **24**: 843–50.
225 Barclay ME, Lyratzopoulos G, Greenberg DC, Abel GA. Missing data and chance variation in public reporting of cancer stage at diagnosis: Cross-sectional analysis of population-based data in England. *Cancer Epidemiol* 2018; **52**: 28–42.
226 Carney M, Quiroga M, Mounce L, Shephard E, Hamilton W, Price S. Effect of pre-existing conditions on bladder cancer stage at diagnosis: a cohort study using electronic primary care records in the UK. *Br J Gen Pract* 2020; **70**: e629–35.
227 Bjerager M, Palshof T, Dahl R, Vedsted P, Olesen F. Delay in diagnosis of lung cancer in general practice. *Br J Gen Pract* 2006; **56**: 863–8.
228 Miller A, Nightingale AL, Sammon CJ, *et al.* Negative Rheumatoid Factor in Primary Care Delays Referral of Patients with Rheumatoid Arthritis. *Rheumatology* 2014; **53**: i29.
229 Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K. False-negative results in screening programmes: Systematic review of impact and implications. *Health Technol Assess (Rockv)* 2000; **4**: 1–120.
230 Yeh MW, Demircan O, Ituarte P, Clark OH. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. *Thyroid* 2004; **14**: 207–15.
231 Montagnana M, Benati M, Danese E. Circulating biomarkers in epithelial ovarian cancer diagnosis: from present to future perspective. *Ann Transl Med* 2017; **5**: 276.
232 Risberg T, Sørbye SW, Norum J, Wist EA. Diagnostic delay causes more psychological distress in female than in male cancer patients. *Anticancer Res* 1996; **16**: 995–9.
233 Norman A, Sisler J, Hack T, Harlos M. Family physicians and cancer care. Palliative care patients' perspective. *Can Fam Physician* 2001; **47**: 2006–16.
234 Skates SJ. Ovarian cancer screening: development of the risk of ovarian cancer algorithm (ROCA) and ROCA screening trials. *Int J Gynecol Cancer* 2012; **22 Suppl 1**: S24–6.
235 Rai N, Champaneria R, Snell K, *et al.* Symptoms, ultrasound imaging and biochemical markers alone or in combination for the diagnosis of ovarian cancer in women with symptoms suspicious of ovarian cancer. *Cochrane Database Syst Rev* 2015; **12**: CD011964.
236 Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: A review and suggested modifications of methodological standards. *J Am Med Assoc* 1997; **277**: 488–94.
237 Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016; **5**: 210.
238 Moons KGM, Wolff RF, Riley RD, *et al.* PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med* 2019; **170**: W1–33.
239 Lurie G, Thompson PJ, McDuffie KE, Carney ME, Goodman MT. Prediagnostic

symptoms of ovarian carcinoma: A case-control study. *Gynecol Oncol* 2009; **114**: 231–6.
240 Rossing MA, Wicklund KG, Cushing-Haugen KL, Weiss NS. Predictive value of symptoms for early detection of ovarian cancer. *J Natl Cancer Inst* 2010; **102**: 222–9.
241 Jordan SJ, Coory MD, Webb PM. Re: Predictive Value of Symptoms for Early Detection of Ovarian Cancer. *JNCI J Natl Cancer Inst* 2010; **102**: 1599–601.
242 Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: Derivation and validation of an algorithm. *Br J Gen Pract* 2013; **63**: e11–21.
243 Collins GS, Altman DG. Identifying women with undetected ovarian cancer: Independent and external validation of QCancer® (Ovarian) prediction model. *Eur J Cancer Care (Engl)* 2013; **22**: 423–9.
244 Grewal K, Hamilton W, Sharp D. Ovarian cancer prediction: Development of a scoring system for primary care. *BJOG An Int J Obstet Gynaecol* 2013; **120**: 1016–9.
245 Kim M-K, Kim K, Kim SM, *et al.* A hospital-based case-control study of identifying ovarian cancer using symptom index. *J Gynecol Oncol* 2009; **20**: 238–42.
246 Macuks R, Baidekalna, Donina S. Diagnostic test for ovarian cancer composed of ovarian cancer symptom index, menopausal status and ovarian cancer antigen CA125. *Eur J Gynaecol Oncol* 2011; **32**: 286–8.
247 Shetty J, Priyadarshini P, Pandey D, Manjunath AP. Modified Goff Symptom Index: Simple triage tool for ovarian malignancy. *Sultan Qaboos Univ Med J* 2015; **15**: e370-5.
248 Jain S, Danodia K, Suneja A, Mehndiratta M, Chawla S. Symptom index for detection of ovarian malignancy in indian women: A hospital-based study. *Journal, Indian Acad Clin Med Indian Acad Clin Med* 2018; **19**: 27–32.
249 Andersen MR, Goff BA, Lowe KA, *et al.* Combining a symptoms index with CA 125 to improve detection of ovarian cancer. *Cancer* 2008; **113**: 484–9.
250 Andersen MR, Goff BA, Lowe KA, *et al.* Use of a Symptom Index, CA125, and HE4 to predict ovarian cancer. *Gynecol Oncol* 2010; **116**: 378–83.
251 Lim AWW, Mesher D, Gentry-Maharaj A, *et al.* Predictive value of symptoms for ovarian cancer: Comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst* 2012; **104**: 114–24.
252 Merritt MA, Green AC, Nagle CM, *et al.* Talcum powder, chronic pelvic inflammation and NSAIDs in relation to risk of epithelial ovarian cancer. *Int J Cancer* 2008; **122**: 170–6.
253 THIN®. The Health Improvement Network. https://www.the-health-improvement-network.com/en/ (accessed Nov 9, 2020).
254 Lowe KA, Shah C, Wallace E, *et al.* Effects of personal characteristics on serum CA125, mesothelin, and HE4 levels in healthy postmenopausal women at high-risk for ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 2480–7.
255 Jacobs I, Prys Davies A, Bridges J, *et al.* Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. *Br Med J* 1993; **306**: 1030–1034.
256 Australian Institute of Health and Welfare, Australasian Association of Cancer Registries, National Cancer Strategies Group. Cancer Incidence Projections Australia 2002 to 2011. Canberra, Australia, 2005.
257 Australian Bureau of Statistics D. Australian Demographic Statistics. Canberra,

Australia, 2007.
258 Garg AX, Adhikari NKJ, McDonald H, *et al.* Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient OutcomesA Systematic Review. *JAMA* 2005; **293**: 1223–38.
259 Chima S, Milley K, Reece JC, Milton S, McIntosh JG, Emery JD. Decision support tools to improve cancer diagnostic decision making in primary care: A systematic review. *Br J Gen Pract* 2019; **69**: e809–18.
260 Chiang PC, Glance D, Walker J, Walter FM, Emery JD. Implementing a qcancer risk tool into general practice consultations: An exploratory study using simulated consultations with Australian general practitioners. *Br J Cancer* 2015; **112**: S77–S83.
261 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; **368**: m441.
262 Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res* 2017; **26**: 796–808.
263 Peduzzi P, Concato J, Kemper E, Holford TR, Feinstem AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**: 1373–9.
264 Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. *Am J Epidemiol* 2007; **165**: 710–8.
265 Riman T, Dickman PW, Nilsson S, *et al.* Risk factors for invasive epithelial ovarian cancer: Results from a Swedish case-control study. *Am J Epidemiol* 2002; **156**: 363–373.
266 Parazzini F, Negri E, Vecchia C La, Restelli C, Franceschi S. Family history of reproductive cancers and ovarian cancer risk: An Italian case-control study. *Am J Epidemiol* 1992; **135**: 35–40.
267 Negri E, Pelucchi C, Franceschi S, *et al.* Family history of cancer and risk of ovarian cancer. *Eur J Cancer* 2003; **39**: 505–10.
268 Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: A systematic review. BJOG. 2005; **112**: 857–65.
269 O'Sullivan JW, Stevens S, Hobbs FDR, *et al.* Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. *BMJ* 2018; **363**: k4666.
270 Ataseven B, Du Bois A, Reinthaller A, *et al.* Pre-operative serum albumin is associated with post-operative complication rate and overall survival in patients with epithelial ovarian cancer undergoing cytoreductive surgery. *Gynecol Oncol* 2015; **138**: 560–5.
271 Merriel SWD, Carroll R, Hamilton F, Hamilton W. Association between unexplained hypoalbuminaemia and new cancer diagnoses in UK primary care patients. *Fam Pract* 2016; **33**: 449–52.
272 Goodman MT, Howe HL, Tung KH, *et al.* Incidence of ovarian cancer by race and ethnicity in the United States, 1992-1997. *Cancer* 2003; **97**: 2676–85.
273 Mathur R, Bhaskaran K, Chaturvedi N, *et al.* Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf)* 2014; **36**: 684–92.
274 World Cancer Research Fund / American Institute for Cancer Research. Diet, nutrition, physical activity and ovarian cancer. 2018. https://www.wcrf.org/sites/default/files/Ovarian-cancer-report.pdf (accessed April

20, 2021).
275 Stewart LM, Spilsbury K, Jordan S, *et al.* Risk of high-grade serous ovarian cancer associated with pelvic inflammatory disease, parity and breast cancer. *Cancer Epidemiol* 2018; **55**: 110–6.
276 Chan JK, Urban R, Capra AM, *et al.* Ovarian cancer rates after hysterectomy with and without salpingo-oophorectomy. *Obstet Gynecol* 2014; **123**: 65–72.
277 Peres LC, Mallen AR, Townsend MK, *et al.* High levels of C-reactive protein are associated with an increased risk of ovarian cancer: Results from the ovarian cancer Cohort Consortium. *Cancer Res* 2019; **15**: 5442–51.
278 Cancer Research UK. Ovarian cancer risk factors. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/risk-factors#heading-Eleven (accessed April 17, 2021).
279 Mounce LTA, Hamilton W, Bailey SER. Cancer incidence following a high-normal platelet count: cohort study using electronic healthcare records from English primary care. *Br J Gen Pract* 2020; **70**: e622–8.
280 Evidence Based Medicine Consult. C-Reactive Protein, CRP (Serum) Level. https://www.ebmconsult.com/articles/lab-test-c-reactive-protein-crp-level (accessed Feb 11, 2021).
281 Baer HJ, Rich-Edwards JW, Colditz GA, Hunter DJ, Willett WC, Michels KB. Adult height, age at attained height, and incidence of breast cancer in premenopausal women. *Int J Cancer* 2006; **119**: 2231–5.
282 The Royal College of Paediatricts and Child Health. Girls UK Growth Chart 2-18 years. 2012. https://www.rcpch.ac.uk/sites/default/files/Girls_2-18_years_growth_chart.pdf (accessed Feb 5, 2020).
283 Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open* 2013; **3**: e003389.
284 Nicholson BD, Aveyard P, Hamilton W, *et al.* The internal validation of weight and weight change coding using weight measurement data within the UK primary care electronic health record. *Clin Epidemiol* 2019; **11**: 145–55.
285 Sterne JAC, White IR, Carlin JB, *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; **338**: b2393.
286 Stata. Impute missing values using chained equations. Stata Manual. https://www.stata.com/manuals13/mimiimputechained.pdf (accessed Dec 14, 2020).
287 Stata. Stata Multiple imputation reference manual (Release 16). 2019. https://www.stata.com/manuals/mi.pdf (accessed Feb 13, 2021).
288 Bodner TE. What Improves with Increased Missing Data Imputations? *Struct Equ Model A Multidiscip J* 2008; **15**: 651–75.
289 White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; **30**: 377–99.
290 Luque-Fernandez MA, Redondo-Sánchez D, Maringe C. cvauroc: Command to compute cross-validated area under the curve for ROC analysis after predictive modeling for binary outcomes. *Stata J* 2019; **19**.
291 Van Calster B, McLernon DJ, van Smeden M, *et al.* Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; **17**: 230.
292 Sperrin M, Marshall AD, Higgins V, Renehan AG, Buchan IE. Body mass index relates weight to height differently in women and older adults: serial cross-sectional surveys

in England (1992–2011). *J Public Heal* 2016; **38**: 607–13.
293 Ensor J. PMSAMPSIZE: Stata module to calculate the minimum sample size required for developing a multivariable prediction model. 2019. https://ideas.repec.org/c/boc/bocode/s458569.html (accessed April 14, 2021).
294 Jacobs I, Oram D, Fairbank J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *BJOG* 1990; **100**: 927–31.
295 Moore RG, McMeekin DS, Brown AK, *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009; **112**: 40–6.
296 Funston G, Hardy V, Abel G, *et al.* Identifying Ovarian Cancer in Symptomatic Women: A Systematic Review of Clinical Tools. *Cancers (Basel)* 2020; **12**: 3686.
297 Nicholson BD, Oke JL, Rose PW, Mant D. Variation in direct access to tests to investigate cancer: A survey of english general practitioners. *PLoS One* 2016; **11**: e0159725.
298 Sundar S, Rick C, Dowling F, *et al.* Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. *BMJ Open* 2016; **6**: e010333.
299 Koo MM, Hamilton W, Walter FM, Rubin GP, Lyratzopoulos G. Symptom Signatures and Diagnostic Timeliness in Cancer Patients: A Review of Current Evidence. *Neoplasia* 2018; **20**: 165–74.
300 Jessen NH, Jensen H, Falborg AZ, Glerup H, Gronbaek H, Vedsted P. Abdominal investigations in the year preceding a diagnosis of abdominal cancer: A register-based cohort study in Denmark. *Cancer Epidemiol* 2021; **72**: 101926.
301 Nicholson BD, Hamilton W, O'Sullivan J, Aveyard P, Hobbs FR. Weight loss as a predictor of cancer in primary care: a systematic review and meta-analysis. *Br J Gen Pract* 2018; **68**: e311–22.
302 Vedsted P, Olesen F. A differentiated approach to referrals from general practice to support early cancer diagnosis - the Danish three-legged strategy. *Br J Cancer* 2015; **112 Suppl**: S65-9.
303 Chapman D, Poirier V, Vulkan D, *et al.* First results from five multidisciplinary diagnostic centre (MDC) projects for non-specific but concerning symptoms, possibly indicative of cancer. *Br J Cancer* 2020; **123**: 722–9.
304 NHS England and NHS Improvement. Rapid Diagnostic Centres Vision and 2019/20 Implementation Specification. London, 2019 https://www.england.nhs.uk/wp-content/uploads/2019/07/rdc-vision-and-1920-implementation-specification.pdf.
305 Kurtz AB, Tsimikas J V., Tempany CMC, *et al.* Diagnosis and Staging of Ovarian Cancer: Comparative Values of Doppler and Conventional US, CT, and MR Imaging Correlated with Surgery and Histopathologic Analysis—Report of the Radiology Diagnostic Oncology Group. *Radiology* 1999; **212**: 19–27.
306 Lee ES, Lee JM. Imaging diagnosis of pancreatic cancer: A state-of-the-art review. *World J Gastroenterol* 2014; **20**: 7864–77.
307 Aberle DR, DeMello S, Berg CD, *et al.* Results of the Two Incidence Screenings in the National Lung Screening Trial. *N Engl J Med* 2013; **369**: 920–31.
308 Suh-Burgmann EJ, Alavi M. Detection of early stage ovarian cancer in a large community cohort. *Cancer Med* 2019; **8**: 7133–40.

309 Richards M. Recovery and Renewal. Report of the Independent Review of Diagnostic Services for NHS England. London, 2020.
310 Rubin G, Walter FM, Emery J, *et al.* Electronic clinical decision support tool for assessing stomach symptoms in primary care (ECASS): a feasibility study. *BMJ Open* 2021; **11**: e041795.
311 Kaiser Permanente. Kaiser Permanente Research Bank. https://researchbank.kaiserpermanente.org/our-research/for-researchers/ (accessed April 15, 2021).
312 The Clinical Practice Research Datalink. Hospital Episode Statistics (HES) Diagnostic Imaging Dataset and CPRD primary care data Documentation (set 18). 2019. https://cprd.com/sites/default/files/Documentation_HES_DID_set18_v1.6.pdf (accessed March 29, 2021).
313 Fonferko-Shadrach B, Lacey AS, Roberts A, *et al.* Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system. *BMJ Open* 2019; **9**: e023232.
314 Walter FM, Thompson MJ, Wellwood I, *et al.* Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. *BMC Cancer* 2019; **19**: 586.
315 Clift AK, Coupland CAC, Hippisley-Cox J. Prostate-specific antigen testing and opportunistic prostate cancer screening: a cohort study in England, 1998–2017. *Br J Gen Pract* 2021; **71**: e157 LP-e165.
316 Luan NN, Wu QJ, Gong TT, Vogtmann E, Wang YL, Lin B. Breastfeeding and ovarian cancer risk: A meta-analysis of epidemiologic studies. *Am J Clin Nutr* 2013; **98**: 1020–1031.
317 Adami HO, Lambe M, Persson I, *et al.* Parity, age at first childbirth, and risk of ovarian cancer. *Lancet* 1994; **344**: 1250–4.
318 Hankinson SE, Colditz GA, Hunter DJ, *et al.* A prospective study of reproductive factors and risk of epithelial ovarian cancer. *Cancer* 1995; **76**: 284–90.
319 Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. *Best Pract Res Clin Obstet Gynaecol* 2017; **41**: 3–14.
320 Shen D, Mao W, Liu T, *et al.* Sedentary behavior and incident cancer: A meta-analysis of prospective studies. *PLoS One* 2014; **9**: e105709.
321 Dixon-Suen SC, Webb PM, Wilson LF, Tuesley K, Stewart LM, Jordan SJ. The association between hysterectomy and ovarian cancer risk: A population-based record-linkage study. *J Natl Cancer Inst* 2019; **111**: 1097–1103.
322 Wentzensen N, Poole EM, Trabert B, *et al.* Ovarian cancer risk factors by histologic subtype: An analysis from the Ovarian Cancer Cohort Consortium. *J Clin Oncol* 2016; **34**: 2888–98.
323 Rasmussen CB, Kjaer SK, Albieri V, *et al.* Pelvic Inflammatory Disease and the Risk of Ovarian Cancer and Borderline Ovarian Tumors: A Pooled Analysis of 13 Case-Control Studies. *Am J Epidemiol* 2017; **185**: 8–20.
324 Barry JA, Azizia MM, Hardiman PJ. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: A systematic review and meta-analysis. *Hum Reprod Update* 2014; **20**: 748–58.
325 Gottschau M, Kjaer SK, Jensen A, Munk C, Mellemkjaer L. Risk of cancer among women with polycystic ovary syndrome: A Danish cohort study. *Gynecol Oncol* 2015; **136**: 99–103.

326 Kelemen LE, Bandera E V., Terry KL, *et al.* Recent alcohol consumption and risk of incident ovarian carcinoma: A pooled analysis of 5,342 cases and 10,358 controls from the Ovarian Cancer Association Consortium. *BMC Cancer* 2013; **13**: 28.
327 Chen HF, Chang YH, Ko MC, Li CY. A large scale population-based cohort study on the risk of ovarian neoplasm in patients with type 2 diabetes mellitus. *Gynecol Oncol* 2014; **134**: 576–80.
328 Weiderpass E, Ye W, Vainio H, Kaaks R, Adami HO. Diabetes mellitus and ovarian cancer (Sweden). *Cancer Causes Control* 2002; **13**: 759–764.
329 Lee J-Y, Jeon I, Kim JW, Song Y-S, Yoon J-M, Park SM. Diabetes Mellitus and Ovarian Cancer Risk: A Systematic Review and Meta-Analysis of Observational Studies. *Int J Gynecol Cancer* 2013; **23**: 402 LP – 412.
330 Wang L, Wang L, Zhang J, Wang B, Liu H. Association between diabetes mellitus and subsequent ovarian cancer in women: A systematic review and meta-analysis of cohort studies. *Medicine* 2017; **96**.
331 Li K, Hüsing a, Fortner RT, *et al.* An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study. *Br J Cancer* 2015; **112**: 1257–65.
332 Prizment AE, Folsom AR, Anderson KE. Nonsteroidal anti-inflammatory drugs and risk for ovarian and endometrial cancers in the Iowa Women's Health Study. *Cancer Epidemiol biomarkers Prev* 2010; **19**: 435–42.
333 Murphy MA, Trabert B, Yang HP, *et al.* Non-steroidal anti-inflammatory drug use and ovarian cancer risk: findings from the NIH-AARP Diet and Health Study and systematic review. *Cancer Causes Control* 2012; **23**: 1839–52.
334 Baandrup L, Faber MT, Christensen J, *et al.* Nonsteroidal anti-inflammatory drugs and risk of ovarian cancer: Systematic review and meta-analysis of observational studies. *Acta Obstet Gynecol Scand* 2013; **92**: 245–55.
335 Havrilesky LJ, Moorman PG, Lowery WJ, *et al.* Oral Contraceptive Pills as Primary Prevention for Ovarian Cancer. *Obstet Gynecol* 2013; **122**: 139–47.
336 Beral V. Ovarian cancer and hormone replacement therapy in the Million Women Study. *Lancet* 2007; **369**: 19–25.
337 Eurostat. Revision of the European Standard Population: Report of Eurostat's task force. Luxembourg, 2013.
338 Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. *PLoS Med* 2020; **17**: e1003295.
339 Goff BA, Lowe KA, Kane JC, Robertson MD, Gaul MA, Andersen MR. Symptom triggered screening for ovarian cancer: A pilot study of feasibility and acceptability. *Gynecol Oncol* 2012; **124**: 230–5.
340 Watson J, Mounce L, Bailey SER, Cooper SL, Hamilton W. Blood markers for cancer. *BMJ* 2019; **367**: l5774.

Appendices

Contents:

Appendix A: Thesis publications

This appendix contains papers (published or under review) arising from this thesis.

Contents:

Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. Garth Funston, Marije Van Melle, Marie-Louise Ladegaard Baun, Henry Jensen, Charles Helsper, Jon Emery, Emma J. Crosbie, Matthew Thompson, Willie Hamilton & Fiona M. Walter. BMC Cancer. 2019; 19:028

The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. Garth Funston, Willie Hamilton, Gary Abel, Emma J. Crosbie, Brian Rous, Fiona M. Walter. PLOS Med. 2020; 17:e1003295

Identifying ovarian cancer in symptomatic women: a systematic review of clinical tools. Garth Funston, Victoria Hardy, Gary Abel, Emma J. Crosbie, Jon Emery, Willie Hamilton, Fiona M. Walter. Cancers. 2020; 8:3686

CA125 test result, test-to-diagnosis interval and stage in ovarian cancer at diagnosis: a cohort study using electronic health records. Garth Funston, Luke Mounce, Sarah Price, Brian Rous, Emma L. Crosbie, Willie Hamilton, Fiona M Walter. Br J Gen Pract. 2021; BJGP.2020.0859 (Online First)

Could ovarian cancer prediction models improve the triage of symptomatic women in primary care? A modelling study using routinely collected data. Garth Funston, Gary Abel, Emma J. Crosbie, Willie Hamilton, Fiona M Walter. (*Under review*)

Funston et al. BMC Cancer (2019) 19:1028
https://doi.org/10.1186/s12885-019-6211-2

BMC Cancer

RESEARCH ARTICLE Open Access

Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines

Garth Funston[1*], Marije Van Melle[1], Marie-Louise Ladegaard Baun[2], Henry Jensen[2], Charles Helsper[3], Jon Emery[4], Emma J. Crosbie[5], Matthew Thompson[6], Willie Hamilton[7] and Fiona M. Walter[1]

Abstract

Background: Women with ovarian cancer can present with a variety of symptoms and signs, and an increasing range of tests are available for their investigation. A number of international guidelines provide advice for the initial assessment of possible ovarian cancer in symptomatic women. We systematically identified and reviewed the consistency and quality of these documents.

Methods: MEDLINE, Embase, guideline-specific databases and professional organisation websites were searched in March 2018 for relevant clinical guidelines, consensus statements and clinical pathways, produced by professional or governmental bodies. Two reviewers independently extracted data and appraised documents using the Appraisal for Guidelines and Research Evaluation 2 (AGREEII) tool.

Results: Eighteen documents from 11 countries in six languages met selection criteria. Methodological quality varied with two guidance documents achieving an AGREEII score ≥ 50% in all six domains and 10 documents scoring ≥50% for "Rigour of development" (range: 7–96%). All guidance documents provided advice on possible symptoms of ovarian cancer, although the number of symptoms included in documents ranged from four to 14 with only one symptom (bloating/abdominal distension/increased abdominal size) appearing in all documents. Fourteen documents provided advice on physical examinations but varied in both the examinations they recommended and the physical signs they included. Fifteen documents provided recommendations on initial investigations. Transabdominal/transvaginal ultrasound and the serum biomarker CA125 were the most widely advocated initial tests. Five distinct testing strategies were identified based on the number of tests and the order of testing advocated: 'single test', 'dual testing', 'sequential testing', 'multiple testing options' and 'no testing'.

Conclusions: Recommendations on the initial assessment and investigation for ovarian cancer in symptomatic women vary considerably between international guidance documents. This variation could contribute to differences in the way symptomatic women are assessed and investigated between countries. Greater research is needed to evaluate the assessment and testing approaches advocated by different guidelines and their impact on ovarian cancer detection.

Keywords: Ovarian cancer, Cancer detection, Ovarian cancer symptoms, Ovarian cancer signs, Ovarian cancer tests, Cancer biomarkers, Symptom-triggered testing, Primary care, Clinical guidelines, Cancer pathways

* Correspondence: gf272@cam.ac.uk
[1]The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
Full list of author information is available at the end of the article

Background

Worldwide, ovarian cancer is the seventh most common cancer in women, with over 200, 000 new cases each year [1]. While once considered a silent killer, it is now recognised that symptoms occur in all stages of disease, although studies differ in the symptoms they report and the positive predictive value (PPV) they attribute to each symptom [2–5]. Given the modest PPVs of individual symptoms, e.g. 0.3% for abdominal pain and 2.5% for abdominal distension, symptoms alone cannot be used to diagnose ovarian cancer, but are routinely used to guide further assessment, including physical examination and testing [4].

An increasing range of tests are used in the initial investigation of symptomatic women for ovarian cancer, including the serum protein biomarker CA125 and imaging modalities such as transabdominal and transvaginal ultrasound, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Algorithms that combine test results with patient characteristics such as age or menopausal state e.g. the Risk of Malignancy Index (RMI) and the ADNEX model, have also been developed to help predict ovarian cancer risk in women presenting with a pelvic mass [6, 7]. However, debate exists regarding the most accurate testing strategy for ovarian cancer. There is very limited research evaluating tests for the initial investigation of symptoms within the primary care setting [8, 9], where most women with this condition first present [10].

Given the discrepancies in the research literature on symptoms and the variety of testing options available, guidance documents, such as clinical practice guidelines, consensus statements and clinical care pathways, have been produced to aid clinicians in making practical decisions regarding the management of women with possible ovarian cancer. As these documents have the potential to significantly affect the care and healthcare outcomes for large numbers of patients, they should be rigorously developed, grounded in the evidence, and make unambiguous recommendations [11, 12].

In this review, we set out to systematically identify and assess the quality of international guidance documents covering the initial assessment for ovarian cancer in symptomatic women. In addition, we aimed to assess the consistency of guidance documents in terms of the symptoms and signs they include and the physical examinations and tests they recommend, to gain an insight into international variation in clinical practice.

Methods

Study selection

We selected documents that provided guidance on the initial assessment of women presenting with symptoms that might represent ovarian cancer i.e. an assessment conducted at the point at which women present with symptoms and enter a given healthcare system. As such, guidance documents that solely provided advice on investigation or management of women after a pelvic mass had been identified, a specialist referral made or a diagnosis of ovarian cancer given, were excluded. As this review focussed on guidance for women presenting with symptoms, the most common mode of ovarian cancer presentation [10, 13], documents which solely provided advice on screening of asymptomatic women or on the investigation of incidental pelvic masses, were excluded. Documents where guidance was limited to sub-groups of patients, e.g. hereditary cancer syndromes, were also excluded. Only documents produced by professional or governmental bodies and published within the ten years before 13th March 2018 were included. There were no language restrictions.

Search strategy

Searches were conducted in Embase and MEDLINE. The MEDLINE search strategy is presented in Additional file 1: Figure S1. Additional searches were performed in guideline specific databases, namely, the National Guideline Clearing House, the Turning Research Into Practice (TRIP) database, the Guidelines International Network, the Canadian Partnership Against Cancer guidelines database, the Canadian Medical Association Infobase and the National Institute of Health and Care Excellence (NICE) website. All searches were performed between 1st and 13th of March 2018. The websites of more than 20 relevant international governmental and professional bodies were hand searched to supplement the database searches.

Guideline selection

Two reviewers independently assessed titles and abstracts. Where either reviewer felt that a document met selection criteria or that it was not possible to exclude on the basis of title and summary alone, the full text was obtained and reviewed against the criteria. Disagreements were resolved by consensus.

Data extraction

Two reviewers, fluent in the language of guideline publication, independently extracted data using a specifically developed template. Discrepancies in extraction were resolved by consensus.

Information on document characteristics (e.g. development body, year of development) and the process of development was collected. We classified documents into one of four categories, which best described their intended purpose and the development process, namely: (1) full Clinical Practice Guidelines (recommendations on patient care, informed by a systematic review of the evidence and taking account of benefits, harms and alternatives) [11]; (2) Short Guides (focused summary

recommendations for patient care, not necessarily based on a full systematic literature review); (3) Consensus Statements (clinically relevant advice based on the opinion of an expert panel) [14], and (4) Clinical Pathways (a structured multidisciplinary plan of patient care, not necessarily based on a full systematic literature review) [15].

The healthcare system for which a guideline is developed will influence the recommendations. We applied a simplified version of the classification system developed by Bohm et al, categorising healthcare systems into three groups: National Health Service, National/Social Health Insurance and Private Health System [16].

Data relating to three components of the initial patient assessment were extracted: symptoms, physical examinations/signs, and investigations. Documents were categorised into the following five groups, based on the number of tests and the order of testing advocated: 'single test' i.e. one test advocated; 'dual testing' i.e. performing two tests concurrently; 'sequential testing' i.e. performing a second type of investigation (second line) if the first type of investigation (first line) is abnormal; 'multiple testing options' i.e. where a range of investigation options were presented with no single investigation being advocated above another; and 'no testing' i.e. where no specific tests were recommended as part of the initial assessment.

Quality assessment

The AGREEII instrument was used to assess the quality of guidance development and reporting of included guidance documents [12]. This validated tool consists of 23 items divided into six domains: 'Scope and Purpose', 'Stakeholder Involvement', 'Rigour of Development', 'Clarity of Presentation', 'Applicability' and 'Editorial Independence'. Each item is rated on a scale from one (criteria not met) to seven (criteria fully met). While developed for clinical practice guidelines, it has been used to assess other types of guidance document [14]. Two reviewers independently assessed each guidance document using the AGREEII tool. Assessments were compared and differences of three or more points per item were discussed and resolved by consensus. Combined scores for each domain were obtained using the following equation: (Obtained score − minimum possible score)/(maximum possible score − minimal possible score) × 100 [12]. We took a score of ≥50% in a particular domain to indicate 'satisfactory' quality [17].

Results

Guideline selection

Our searches identified 846 documents, of which 178 were duplicates. The titles and summaries of 668 documents were screened, and 62 full text documents were obtained for further scrutiny. Eighteen documents met our selection criteria (Fig. 1).

Guideline characteristics

Of the 18 documents that met the selection criteria, two were developed in continental Europe, five in the United Kingdom (UK) and Republic of Ireland, three in Scandinavia, four in North America and four in Australasia (Table 1) [18, 21–37]. Thirteen documents were published in English. Ten documents were categorised as full clinical practice guidelines, three as short guides, four as clinical pathways and one as a consensus statement. Documents varied in their intended audience and scope. Some dealt only with the initial assessment and referral of symptomatic patients and were aimed primarily at primary care practitioners [24, 26, 32–34]. Others also dealt with definitive diagnosis and treatment, often devoting more attention to this than initial assessment, and appeared to have a broader target audience including primary care practitioners and specialists [21, 22, 25, 29, 31, 35, 36]. Nine documents were developed for countries with National/Social Health Insurance Systems, seven for countries with National Health Services and two for a country with a Private Healthcare System.

Quality assessment

Two guidance documents scored ≥50% in all six domains (Additional file 1: Table S1). Scores for the Rigour of Development domain (which appraises the process of evidence identification, synthesis, assessment and recommendation formulation) ranged from 7 to 96%, with 10 documents scoring ≥50% (Table 1).

Symptoms

All guidance documents provided advice regarding presenting symptoms that should prompt a doctor to consider ovarian cancer. The numbers of guidelines in which each symptom was included is shown in Fig. 2. One or more of the related terms bloating, abdominal distention, increased abdominal size or girth, were listed as symptoms of ovarian cancer in all documents, abdominal or pelvic pain in 16 documents, urinary frequency in 14 documents and feeling full or early satiety in 14 documents. We identified 20 symptom terms that were included in under 50% of documents. The number of symptom terms included in the recommendations of documents ranged from four to 14 (Additional file 1: Table S2). Some documents simply listed symptoms doctors should be aware of in relation to ovarian cancer, while others provided further details on symptom frequency (e.g. > 12x/month), nature (e.g. persistent), duration (e.g. > 1 year) and age at presentation (e.g. > 50 years).

Physical examinations and signs

Fourteen documents provided guidance on physical examination or the signs associated with ovarian cancer (Table 2). Thirteen of these documents specifically

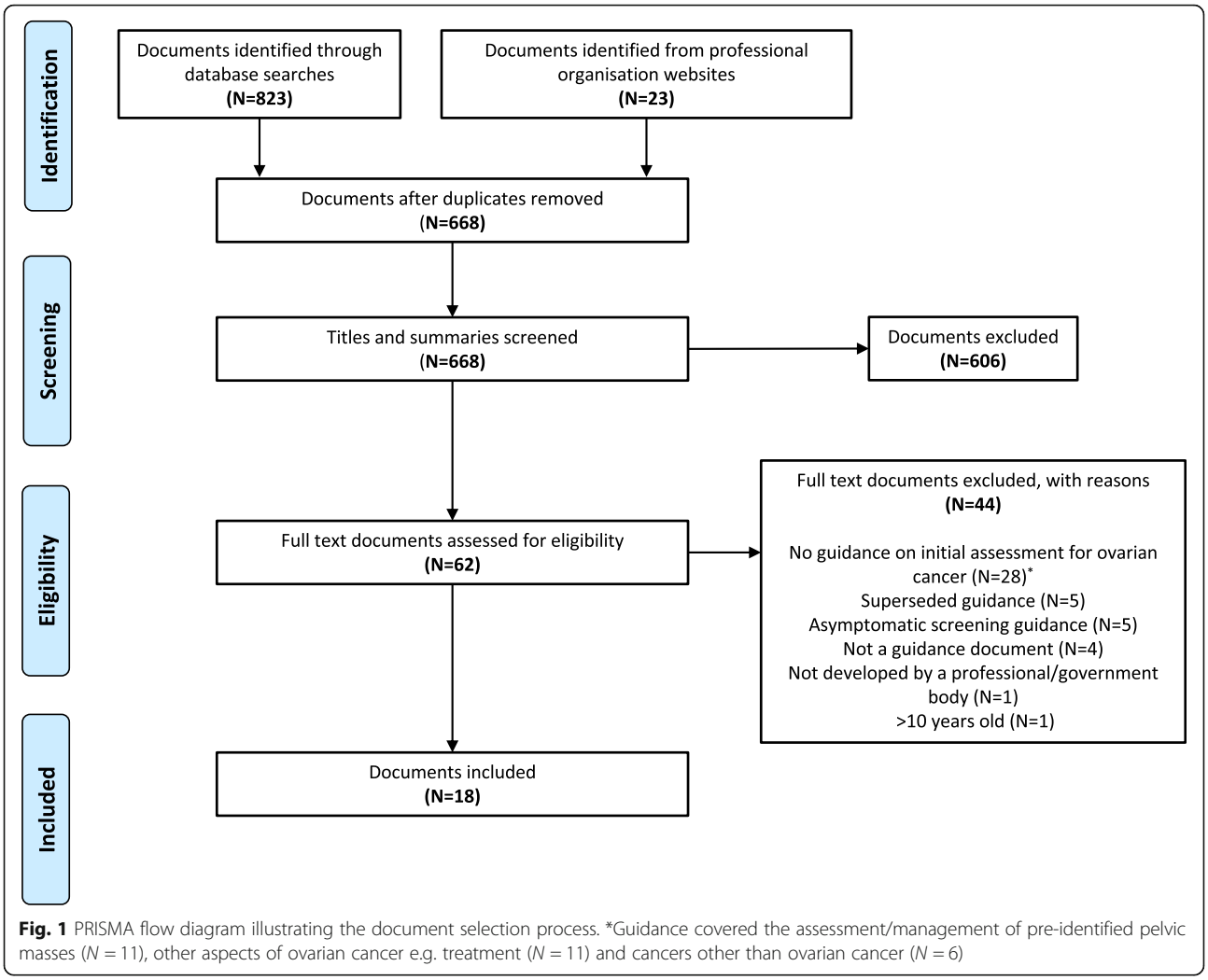


Fig. 1 PRISMA flow diagram illustrating the document selection process. *Guidance covered the assessment/management of pre-identified pelvic masses (*N* = 11), other aspects of ovarian cancer e.g. treatment (*N* = 11) and cancers other than ovarian cancer (*N* = 6)

advocated abdominal examination or mentioned abdominal signs. Nine documents specifically advocated pelvic or gynaecological examination, three of which detailed that this should include a speculum examination, three a bimanual or digital examination and one a vaginal examination, while three documents recommended a rectal examination.

Tests

Fifteen documents provided advice on the initial investigation of symptoms and were categorised based on the number and order of tests recommended (Table 3). One document advocated a single test strategy, four a duel testing strategy, four a sequential testing strategy, three gave multiple testing options, and three did not advocate testing prior to referral, although two of these did recommend that a CA125 sample be taken at the point of specialist referral so as to be available to the specialist. One document could not be categorised as it was unclear when and how tests should be used in the initial assessment for ovarian cancer [21]. The most commonly advocated tests for initial investigation were CA125 (11 documents) and ultrasound (12 documents). Several guidelines also recommended using additional cancer biomarkers such as CA19–9, CEA, AFP and HCG, routine blood tests including full blood count and renal function, imaging tests including CT and MRI, and the risk tools RMI and ADNEX.

Although the majority of guidelines used symptoms as the trigger for initiating tests, the two Australian short guides indicated that testing for ovarian cancer should be conducted if there was a suspicion on clinical examination [23, 24]. Conversely, guidelines from Ireland, England, Scotland, the UK, Sweden and Norway recommended that concerning findings on examination should prompt an urgent referral to a specialist rather than tests [18, 31–34, 37].

Discussion

In the absence of effective screening programmes, most women are diagnosed with ovarian cancer following the

Table 1 Characteristics of guidance documents presented by geographical area

Guidance document	Development body	Publication date of current version	Country and language if other than English	CPG	SG	CP	CS	Rigour of development (AGREEII) %	Healthcare system
Continental Europe									
Epithelial ovarian carcinoma	Dutch Society for Obstetrics and Gynaecology (NVOG)	2018	Netherlands (Dutch)	♦				66	National/ Social Health Insurance
Guideline on diagnostics, therapy and follow-up of malignant ovarian tumours	The Association of Scientific Medical Societies in Germany (AWMF), led by German Society for Gynaecology and Obstetrics (DGGG)	2017	Germany (German)	♦				81	National/ Social Health Insurance
United Kingdom and Republic of Ireland									
Epithelial ovarian / fallopian tube / primary peritoneal cancer guidelines: recommendations for practice	British Gynaecological Cancer Society	2017	UK	♦				48	National Health Service
Ovarian cancer GP referral for symptomatic women	National Cancer Control Programme	2016	Republic of Ireland		♦			7	National/ Social Health Insurance
Suspected cancer: recognition and referral	National Institute for Health and Care Excellence (NICE)	2015	England, Wales, Northern Ireland	♦				96	National Health Service
Scottish referral guidelines for suspected cancer	Healthcare Improvement Scotland	2014	Scotland	♦				55	National Health Service
Management of epithelial ovarian cancer	Scottish Intercollegiate Guidelines Network (Part of Healthcare Improvement Scotland)	2013	Scotland	♦				76	National Health Service
Scandinavia									
Integrated ovarian cancer patient pathway	The Danish National Health Authority	2016	Denmark (Danish)			♦		29	National Health Service
Ovarian cancer patient pathway	The Norwegian Directorate of Health	2016	Norway (Norwegian)			♦		38	National Health Service
Standardised ovarian cancer care pathway [a]	Regional Cancer Centre Co-operative Sweden	2015	Sweden (Swedish)	♦				55	National Health Service
Australasia									
Assessment of symptoms that may be ovarian cancer: a guide for general practitioners[b]	Cancer Australia	2015	Australia		♦			50	National/ Social Health Insurance
Appropriate referral of women with suspected ovarian cancer[b]	Cancer Australia	2015	Australia		♦			50	National/ Social Health Insurance
Optimal care pathway for women with ovarian cancer	Cancer Council Victoria	2015	Australia			♦		10	National/ Social Health Insurance

Table 1 Characteristics of guidance documents presented by geographical area *(Continued)*

Guidance document	Development body	Publication date of current version	Country and language if other than English	CPG	SG	CP	CS	Rigour of development (AGREEII) %	Healthcare system
Suspected cancer in primary care: Guidelines for investigation, referral and reducing ethnic disparity	New Zealand Guidelines Group	2009	New Zealand	♦				56	National/ Social Health Insurance
North America									
Ovarian cancer: including fallopian tube cancer and primary peritoneal cancer	National Comprehensive Cancer Network	2018 (v2)	USA	♦				65	Private Health System
The role of the obstetrician-gynaecologist in the early detection of epithelial ovarian cancer in women at average risk	American College of Obstetrician Gynaecologists and the Society of Gynaecological Oncology	2017	USA				♦	11	Private Health System
Ovarian cancer diagnosis pathway map	Cancer Care Ontario	2016	Ontario, Canada			♦		19	National/ Social Health Insurance
Genital tract cancers in females: ovarian, fallopian tube, and primary peritoneal cancers	Guidelines and Protocol Advisory Committee (Medical Services Commission)	2014	British Columbia, Canada	♦				16	National/ Social Health Insurance

CPG Clinical Practice Guideline, *SG* Short Guideline, *CP* Clinical Pathway, *CS* Consensus Statement

[a]A full clinical practice guideline covering initial assessment, definitive diagnosis and treatment [18], and a short version focussing on initial assessment and investigation in primary care [19], are available. Guidance on initial assessment differed slightly between the two documents. The recommendations presented in this review were extracted from the short guide. AGREEII appraisal included an assessment of the full guideline evidence review

[b]Short guide, still active. Based on a now rescinded 2004 full clinical practice guideline entitled 'Clinical practice guidelines for the management of women with ovarian cancer' [20]. AGREEII appraisal included an assessment of the full guideline evidence review

onset of symptoms [10, 13]. In this review, we identified and compared international guidance documents on the initial assessment and investigation for possible ovarian cancer in symptomatic women. Our results highlight significant differences between international guidelines, not only in the clinical features they suggest should trigger a suspicion of ovarian cancer, but also in the initial examinations and investigations they advocate.

The stage distribution of ovarian cancer at diagnosis, and ovarian cancer survival, varies between countries [38]. A positive correlation has been demonstrated between national survival and the readiness of primary care practitioners to investigate or refer women with symptoms of possible ovarian cancer [39]. International variation in the way symptomatic women are assessed and investigated could also contribute to differences in the timeliness of ovarian cancer diagnosis and survival. Although guidelines are not always followed [40], they do influence practice [41, 42], and variation in international guidelines is likely to indicate differences in clinical practice internationally. International comparative research is ongoing to investigate differences in access to tests for ovarian cancer and survival [43]. Several studies have sought to evaluate the impact of national urgent cancer referral guidelines on timeliness of diagnosis and/or survival [42, 44, 45], but there is little research similarly evaluating the effect of guidelines which advocate symptom-triggered testing for ovarian cancer [46]. Studies are needed to evaluate the impact of such guidance to ensure that the recommended approaches are effective, for example, by comparing stage distribution and cancer survival pre- and post- implementation of guidance. Comparing the impact of cancer detection guidelines between countries is challenging, not least as it relies on the use of standardised endpoints (stage, survival) which are not always uniformly recorded. Initiatives such as the International Cancer Benchmarking Partnership [43], may improve consistency in the recording of such outcomes and so aid international comparisons.

Guideline developers have to consider the healthcare system for which they are developing guidance. The guidance from countries with National Health Services was, in general, specific on symptoms and signs and gave clear recommendations on which tests should be performed and in what order. In contrast, guidance from the USA, which has a Private Healthcare System, was much less prescriptive, providing different options for the clinician. This is likely to reflect the fact that National Health Services aim to provide uniform services and level of care across a country/region and must plan for this, while the care provided in a country with a Private Healthcare System may differ depending on the private provider. Similarly, guideline recommendations

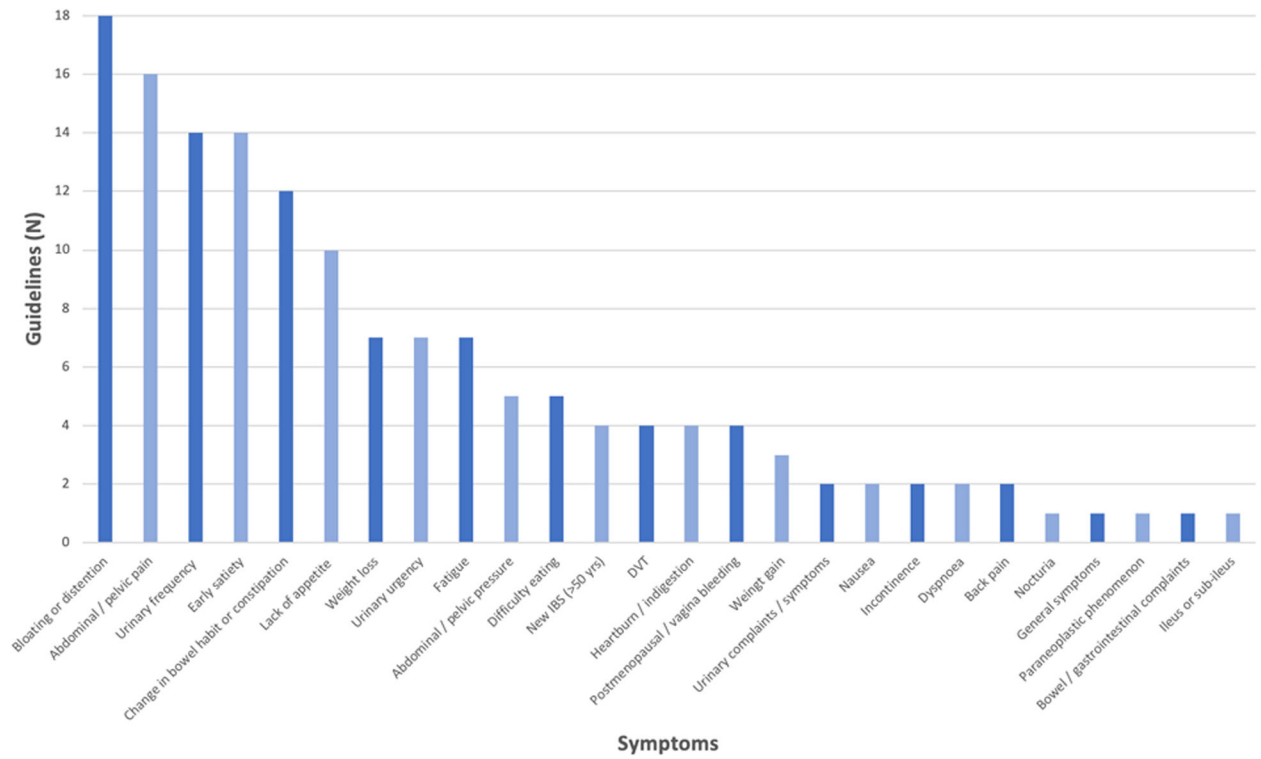


Fig. 2 Symptoms included in guidelines

may be influenced by the speciality of the clinician performing the initial assessment within a healthcare system e.g. GP/family physician and/or gynaecologist. Gynaecologists may be more competent with, and willing to perform, gynaecological examinations and better equipped to interpret complex tests and algorithms. Direct access to gynaecologists is available in the USA and Germany and guidance from these countries included a range of specialist tests [47, 48]. In contrast, in countries like the UK, Ireland, Australia and Scandinavia, where GPs play a strong gatekeeping role and where a referral is generally required prior to gynaecology assessment, a limited number of tests were recommended.

Over the last 15 years a number of studies have explored associations between ovarian cancer and symptoms; however, differences exist between the symptoms they have identified and their predictive values. Most documents in this review included symptoms widely regarded as increasing the likelihood of an ovarian cancer being present, for example, abdominal distension and pelvic pain [4, 5, 49]. Some documents also included symptoms such as fatigue, nausea, back pain and the generic term 'urinary symptoms', which are more controversial, and were not found to increase the likelihood of ovarian cancer in a recent comprehensive systematic review [49]. Some variation may be due to the type of evidence that guideline developers chose to consider. For example, UK guideline developers appear to have taken account of all relevant international studies when deciding which symptoms should be included in the guidance [8]. In contrast, USA guidelines included a more restricted list of symptoms derived from the influential Ovarian Cancer Symptom Index which was developed in the USA [50]. As almost all published studies exploring associations between ovarian cancer and symptoms have been undertaken in the UK and the USA, guideline developers outside these countries must rely on international evidence to inform their recommendations [49]. Further large, high quality research studies, undertaken in countries around the world, would improve our understanding of the symptomology of ovarian cancer and help resolve disagreements over which symptoms should be included in guidelines.

Given the range of AGREEII scores guidelines obtained in the Rigour of Development domain, discrepancies in symptoms and other recommendations are likely stem in part from differences in the scope and quality of evidence reviews undertaken by guideline developers. It is likely that where a rigorous systematic approach is not followed, important research, for example on symptoms, may be missed. All guidance documents in this review are likely to influence patient care and should be developed rigorously and be explicit about the development process. Different strategies could help encourage this, which in

Table 2 Physical examinations recommended and ovarian cancer signs noted within guidance documents

Document	Type of examination specified	Signs
Continental Europe		
Epithelial ovarian carcinoma (Neth)	Not specified	- Pelvic mass / abdominal mass
		- Ascites
		- Pleural effusion
		- Increased uterine / vaginal prolapse
		- Enlarged supraclavicular lymph nodes
Guideline on diagnostics, therapy and follow-up of malignant ovarian tumours (Ger)	Abdominal and pelvic / gynaecological examination (including digital and speculum)	- Ovarian mass
United Kingdom and the Republic of Ireland		
Epithelial Ovarian / Fallopian Tube / Primary Peritoneal Cancer Guidelines: recommendations for practice (UK)	Examination	- Pelvic or abdominal mass
Suspected cancer: recognition and referral (Eng)	Physical examination	- Ascites
		- Pelvic / abdominal mass (not obviously uterine fibroids)
Ovarian cancer GP referral for symptomatic women (Ire)	Clinical examination (include a bimanual-pelvic examination)	- Unexplained ascites
		- Pelvic mass
		- Palpable ovaries in postmenopausal women
Scottish referral guidelines for suspected cancer (Scot)[a]	Abdominal palpation	- Ascites
		- Pelvic or abdominal mass (not obviously uterine fibroids, gastrointestinal or urological in origin)
Management of epithelial ovarian cancer (Scot)	Not specified	- Not specified
Scandinavia		
Integrated ovarian cancer patient pathway (Den)	Gynaecological examination (including palpation and speculum)	- Ascites
		- Pelvic mass
Ovarian cancer patient pathway (Nor)	Not specified	- Not specified
Standardised ovarian cancer care pathway (Swed)[b]	Palpation of superficial lymph nodes, abdominal palpation, rectal examination and auscultation of the heart and lungs	- Pleural effusion (unexplained)
		- Ascites
Australasia		
Assessment of symptoms that may be ovarian cancer: a guide for general practitioners (Aus)	Abdominal palpation, pelvic assessment, vaginal and rectal examination	- Firm resistance on abdominal palpation
		- Unexplained fullness
		-Fullness + shifting dullness on percussion
		- Hard irregular mass in the pouch of Douglass
		- Adnexal mass
Appropriate referral of women with suspected ovarian cancer (Aus)	Not specified	- Not specified
Optimal care pathway for women with ovarian cancer (Aus)	General and pelvic examination	- Not specified
Suspected cancer in primary care: guidelines for investigation, referral and reducing ethnic disparity (NZ)	Abdominal palpation and pelvic examination	- Not specified
North America		
Ovarian cancer: including fallopian tube cancer and primary peritoneal cancer (USA)	Abdominal and pelvic examination	- Suspicious palpable pelvic or abdominal mass
		- Ascites or abdominal distension

Table 2 Physical examinations recommended and ovarian cancer signs noted within guidance documents *(Continued)*

Document	Type of examination specified	Signs
The role of the obstetrician-gynaecologist in the early detection of epithelial ovarian cancer in women at average risk (USA)	Not specified	- Not specified
Ovarian cancer diagnosis pathway map (Ont, Can)	Directed physical examination. Pelvic examination including speculum and bimanual examinations and examination of the external genitalia	- Suspicious palpable pelvic or abdominal mass - Ascites
Genital tract cancers in females: ovarian, fallopian tube, and primary peritoneal cancers (BC, Can)	A physical examination of the abdomen and pelvis including a pelvi-rectal examination	- Abdominal mass

[a]As recorded on associated Microsite and Short guidance document. The full guideline covers all gynaecological cancers with examinations and findings listed together. Microsite and Short guideline lists examinations and signs by cancer site
[b]Both a full clinical practice guideline covering initial assessment, definitive diagnosis and treatment, and a short version focusing on initial assessment and investigation in primary care, are available. Guidance on initial assessment differed slightly between the two documents. The presented data was extracted from the short guide

turn could help to harmonise symptoms in international guidelines. For example, funders could have guidelines independently appraised following development, using the AGREEII checklist, and publish the results alongside the guidelines. In addition, many guidelines are published in peer reviewed journals. Guideline developers could be required to submit an AGREEII style checklist as part of the submission process. While not all guideline development groups have the significant resources required to develop all elements of clinical guidelines de novo, this may not be necessary. For example, the guidance from the New Zealand Guideline Group was based on 2005 NICE guidance and adapted to suit the New Zealand healthcare system. Collaboration by international guideline producers on aspects of guidelines such as symptoms, which are likely to differ little between healthcare systems or countries, could also help reduce duplication, ensure quality and increase consistency.

A pelvic or gynaecological examination was specifically recommended by half of the guidelines, with three specifying that a speculum and three a bimanual or digital examination, be performed. However, Myres et al.'s review, which included studies on examinations performed by gynaecologists pre-surgery and in the screening setting, found that less than half of adnexal masses are picked up on bimanual examination [51]. GPs might be less skilled at identifying pelvic masses, but a recent review identified no studies evaluating their competence at performing pelvic examinations for gynaecological cancer [52].

Most documents recommended the use of ultrasound and/or CA125 in the initial investigation for ovarian cancer. However, guidelines varied in the sequence of testing, and a variety of other serum biomarkers, imaging modalities and risk algorithms were included in some. This variation may result in part from differences in the funding and available resources within different healthcare systems. For example, consideration of costs and resource implications played a role in the decision by NICE to recommend the relatively cheap and widely accessible CA125 test rather than ultrasound as the first line investigation [8]. There is little high quality evidence for tests used in the initial investigation of possible ovarian cancer [8], often necessitating consensus opinion [34, 35], with one guideline making no recommendations on testing because of the lack of evidence [26].

Evidence from secondary care and screening studies indicates that CA125 and ultrasound differ in their diagnostic accuracy [8, 53, 54]. Therefore, the test(s) chosen, and, where they are used in combination, the order of testing, may have important implications for cancer detection. For example, a sequential testing approach, where both tests need to be abnormal to trigger specialist referral [33], will be more specific at the cost of lower sensitivity. Conversely, a dual-testing approach, where an abnormality in either test warrants referral [34, 35], will be more sensitive but sacrifices specificity and economy.

This is the first study to systematically identify and compare international guidance documents on the initial assessment and investigation for possible ovarian cancer in symptomatic women. Direct comparisons between the testing strategies employed in different countries must be interpreted with reference to the healthcare system for which the guidance was produced. Although we performed a comprehensive literature search, it is possible that we did not identify all relevant guidance documents e.g. healthcare guidelines not published online or not available outside the region or country of publication. We attempted to obtain all relevant documentation on the development process of guidelines included in this review, contacting guideline producers for additional information when necessary, to allow us to perform comprehensive AGREEII appraisals. However, it is possible that we did not gain access to all relevant documents e.g. unpublished search strategies or evidence reviews.

Table 3 Summary of tests recommended for the assessment of symptoms and/or signs of ovarian cancer

Strategy	Guideline	When is testing advocated?	Initial tests
Single test	Guideline on diagnostics, therapy and follow-up of malignant ovarian tumours (Ger)	Signs or symptoms of ovarian cancer (OC)	Transvaginal US *Note: CT, MRI, PET CT may be used in specific cases*
Dual testing	Scottish referral guidelines for suspected cancer (Scot)	Symptoms of OC *Note: Ascites- refer urgently rather than test*	CA125 + pelvic US
	Management of epithelial ovarian cancer (Scot)	Symptoms of OC	CA125 + pelvic US
	Assessment of symptoms that may be ovarian cancer: a guide for general practitioners (Aus)	Mass identified clinically *Note: No mass identified clinically- refer appropriately*	CA125 + transvaginal US *Or* CA125 + Abdominal US *Or* CA125 + CT
	Appropriate referral of women with suspected ovarian cancer (Aus)	Suspicious findings on clinical examination	CA125 + transvaginal US +/− calculation of Risk of Malignancy Index (RMI)
Sequential testing	Suspected cancer: recognition and referral (Eng)	OC symptoms *Note: Ascites or suspicious mass- refer urgently rather than test*	First line: CA125 Second line: Abdominopelvic US (if CA125 is abnormal)
	Epithelial ovarian / fallopian tube / primary peritoneal cancer guidelines: recommendations for practice (UK)	OC symptoms *Note: Pelvic or abdominal mass- refer urgently rather than test*	First line: CA125 Second line: Abdominopelvic US (if CA125 is abnormal)
	Ovarian cancer GP referral for symptomatic women (Ire)	History suspicious of OC but examination normal *Note: Suspicious pelvis mass or ascites- refer urgently rather than test*	First line: CA125 Second line: US of pelvis (If CA125 35–200 u/ml) *Note:* If CA125 > 200 u/ml refer without US
	Ovarian cancer diagnosis pathway map (Ont, Can)	Suspicion of OC *Note: Tests may be performed prior to specialist referral but are not a requirement for referral. Can refer prior to testing*	First line: Transvaginal US and / or other imaging Second line: CA125, FBC, Renal Function + RMI *(*If indicated: *CEA, CA19–9, other tumour markers* e.g. *AFP, LDH, HCG)*
Multiple testing options	Optimal care pathway for women with ovarian cancer (Aus)	Symptoms of OC	Pelvic US + Routine blood tests + CA125 + Algorithms such as RMI, ADNEX +/− CT scan
	Genital tract cancers in females: ovarian, fallopian tube, and primary peritoneal cancers (BC, Can)	Suspicion of OC *Note: Imaging not essential for referral*	Transvaginal or abdominal US Blood tests: CA125**,** CA19–9**,** CA15–3, CEA *< 40 yrs old:* AFP, HCG, LDH
	Ovarian cancer Including fallopian tube cancer and primary peritoneal cancer (USA)	Suspicion of OC *Note: Provides some advice on when particular tests are indicated. Appears to include both initial and pre-surgical tests*	US *and/or* abdominal/pelvic CT/MRI (as indicated) Chest CT or chest x-ray (as indicated) Complete blood count, chemistry profile and LFT CA125 or other tumour markers (as indicated: inhibin, β-hCG, AFP, LDH, CEA, CA19–9) Nutritional status GI evaluation (as indicated)
No testing prior to referral	Integrated ovarian cancer patient pathway (Den)	At point of specialist referral	*Note CA125 requested in primary care at time of referral so as to be available to the specialist. Not acted upon in primary care*
	Ovarian cancer patient pathway (Nor)	Post specialist referral	Post referral
	Standardised ovarian cancer care pathway (Swed)	At point of specialist referral	*Note CA125 requested in primary care at time of referral so as to be available to the specialist. Not acted upon in primary care*
Unclear or no recommendations on testing given	Suspected cancer in primary care: guidelines for investigation, referral and reducing ethnic disparity (NZ)	No recommendations	No recommendations

Table 3 Summary of tests recommended for the assessment of symptoms and/or signs of ovarian cancer *(Continued)*

Strategy	Guideline	When is testing advocated?	Initial tests
	The role of the obstetrician-gynaecologist in the early detection of epithelial ovarian cancer in women at average risk (USA)	No recommendations	No recommendations
	Epithelial ovarian carcinoma (Netherlands)	Suspicion of OC. Not clear which tests should be used and when they should be used for initial investigation	Blood tests discussed: routine blood tests, CA125 +/− CEA

Guidelines are grouped into categories on the bases of the number and order of tests advocated

Conclusion

Multiple international guidance documents provide advice on the initial assessment and investigation for possible ovarian cancer in symptomatic women. These documents differ markedly in the symptoms they include and the physical examinations and clinical investigations they recommend. Given this, it is probable that patient care and the likelihood of cancer detection will vary depending on the guidance document followed. Studies evaluating the role of examinations and the diagnostic performance of testing strategies for the initial assessment of possible ovarian cancer in symptomatic women are needed to aid the development of more evidence-based guidelines.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12885-019-6211-2.

Additional file 1: Figure S1. Medline search strategy. **Table S1.** Scores in percent for each domain of guidance documents calculated using the AGREEII tool. **Table S2.** Summary of symptoms included in each guidance document. (DOCX 26 kb)

Abbreviations

AGREEII: Appraisal for Guidelines and Research Evaluation 2; CT: Computed Tomography; MRI: Magnetic Resonance Imaging; NICE: National Institute of Health and Care Excellence; PPV: Positive Predictive Value; RMI: Risk of Malignancy Index; TRIP: Turning Research Into Practice; UK: United Kingdom

Acknowledgements

Not applicable.

Authors contributions

GF designed the study and performed the searches. GF and MVM screened the titles and summaries. GF, MVM, MLLB, HJ and CH selected full text documents for inclusion and extracted data. GF and FMW interpreted the data. GF wrote the paper. MVM, MLLB, HJ, CH, JE, EJC, MT, WH and FMW reviewed and commented on the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Cancer Research UK [grant number: C8640/A23385]. Cancer Research UK had no further role in the study design; in the collection, analysis and interpretation of data; in the writing of the article or in the decision to submit the article for publication.

Availability of data and materials

This study was based entirely on previously published data which is available online from the sources described in the article. No datasets were developed or analysed in this study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

WH was clinical lead for the 2015 NICE guidelines 'Suspected cancer: recognition and referral' (NG12), which was included in this review. WH did not take part in the AGREEII assessment of guidelines for this review. WH contributed to this article in a personal capacity and his contribution should not be interpreted as representing the views of NICE or the guideline development group. All other authors declare no conflict of interest.

Author details

[1]The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [2]Research Centre for Cancer Diagnosis in Primary Care, Research Unit for General Practice, Aarhus University, Aarhus, Denmark. [3]Julius Centre for Health Sciences and Primary Care, Utrecht University, Utrecht, Netherlands. [4]Centre for Cancer Research and Department of General Practice, University of Melbourne, Melbourne, Australia. [5]Gynaecological Oncology Research Group, Division of Cancer Sciences, University of Manchester, Manchester, UK. [6]Department of Family Medicine, University of Washington, Seattle, USA. [7]University of Exeter Medical School, University of Exeter, Exeter, UK.

Received: 27 January 2019 Accepted: 26 September 2019
Published online: 01 November 2019

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global Cancer statistics, 2012. CA Cancer J Clin. 2015;65:87–108.
2. Goff BA, Mandel L, Muntz HG, Melancon CH. Ovarian carcinoma diagnosis. Cancer. 2000;89:2068–75.
3. Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. BJOG. 2005;112:857–65.
4. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. BMJ. 2009;339:b2998.
5. Goff BA, Mandel LS, Melancon CH, Muntz HG. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. JAMA. 2004;291:2705–12.
6. Jacobs I, Oram D, Fairbank J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. BJOG. 1990;100:927–31.
7. Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. BMJ. 2014;349:g5920.
8. National Institute for Health and Clinical Excellence. Ovarian Cancer: the recognition and initial management of ovarian cancer. Cardiff, UK: National Collaborating Centre for Cancer, 2011.
9. Sundar S, Neal RD, Kehoe S. Diagnosis of ovarian cancer. BMJ. 2015;351:h4443.
10. Barrett J, Sharp DJ, Stapley S, Stabb C, Hamilton W. Pathways to the diagnosis of ovarian cancer in the UK: a cohort study in primary care. BJOG. 2010;117:610–4.
11. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. Clinical practice guidelines we can trust. Washington DC, United States: National Academies Press, 2011.

12. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010;182:E839–42.
13. Nagle CM, Francis JE, Nelson AE, Zorbas H, Luxford K, de Fazio A, et al. Reducing time to diagnosis does not improve outcomes for women with symptomatic ovarian Cancer: a report from the Australian ovarian Cancer study group. J Clin Oncol. 2011;29:2253–8.
14. Nagler EV, Vanmassenhove J, van der Veer SN, Nistor I, Van Biesen W, Webster AC, et al. Diagnosis and treatment of hyponatremia: a systematic review of clinical practice guidelines and consensus statements. BMC Med. 2014;12:1.
15. Lawal AK, Rotter T, Kinsman L, Machotta A, Ronellenfitsch U, Scott SD, et al. What is a clinical pathway? Refinement of an operational definition to identify clinical pathway studies for a Cochrane systematic review. BMC Med. 2016;14:35.
16. Böhm K, Schmid A, Götze R, Landwehr C, Rothgang H. Five types of OECD healthcare systems: empirical results of a deductive classification. Health Policy (New York). 2013;113:258–69.
17. Huang T-W, Lai J-H, Wu M-Y, Chen S-L, Wu C-H, Tam K-W. Systematic review of clinical practice guidelines in the diagnosis and management of thyroid nodules and cancer. BMC Med. 2013;11:191.
18. Regional Cancer Centre Co-operative Sweden. Standardised Ovarian Cancer Care Pathway. 2015. https://www.cancercentrum.se/globalassets/cancerdiagnoser/gynekologi/aggstockscancer/vardforlopp/standardiserat_vardforlopp_aggstockar_20151221.pdf. Accessed 25 Aug 2018.
19. Regional Cancer Centre Co-operative Sweden. Standardised ovarian cancer care pathway, abreviated version for primary care. 2015. Available at: https://www.cancercentrum.se/globalassets/vara-uppdrag/kunskapsstyrning/varje-dag-raknas/vardforlopp/kortversioner/pdf/kortversion-svf-aggstockscancer.pdf. Accessed: 29th Aug 2018.
20. Australian Cancer Network and the National Breast Cancer Centre. Clinical practice guidelines for the management of women with epithelial ovarian cancer. 2004. Available at: https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/clinical-practice-guidelines-management-women-epithelial-ovarian-cancer. Accessed: 6th Aug 2018.
21. Dutch Society For Obstetrics and Gynaecology. Epithelial ovarian carcinoma. 2018. www.oncoline.nl/ovariumcarcinoom. Accessed 12 Jul 2018.
22. German Guideline Programme in Oncology (OL) of the Association of the Scientific Medical Societies in Germany (AWMF). Guideline on diagnostics, therapy and follow-up of malignant ovarian tumours. 2017. https://www.awmf.org/leitlinien/detail/ll/032-035OL.html. Accessed 4 Aug 2018.
23. Cancer Australia. Assessment of symptoms that may be ovarian cancer: a guide for general practitioners. 2015. https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/assessment-symptoms-may-be-ovarian-cancer-guide-gps. Accessed 12 Jul 2018.
24. Cancer Australia. Appropriate referral of women with suspected ovarian cancer. 2015. https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/appropriate-referral-women-suspected-ovarian-cancer. Accessed 12 Jul 2018.
25. Cancer Council Victoria. Optimal care pathway for women with ovarian cancer. 2015. https://www.cancervic.org.au/downloads/health-professionals/optimal-care-pathways/Optimal_care_pathway_for_women_with_ovarian_cancer.pdf. Accessed 31 Aug 2018.
26. New Zealand Guidelines Group. Suspected Cancer in primary care: guidelines for investigation, referral and reducing ethnic disparity. 2009. https://www.health.govt.nz/publication/suspected-cancer-primary-care-guidelines-investigation-referral-and-reducing-ethnic-disparities. Accessed 12 Jul 2018.
27. Society of Gynecologic Oncology. Committee Opinion No. 716: The Role of the Obstetrician-Gynecologist in the Early Detection of Epithelial Ovarian Cancer in Women at Average Risk. Obstet Gynecol. 2017;130:e146–9.
28. Cancer Care Ontario. Ovarian cancer diagnosis pathway map. 2016. www.cancercareontario.ca/en/pathway-maps/ovarian-cancer. Accessed 12 Jul 2018.
29. National Comprehensive Caner Network. Ovarian cancer: including fallopian tube cancer and primary peritoneal cancer. 2018. https://www.nccn.org/professionals/physician_gls/default.aspx#ovarian. Accessed 31 Aug 2018.
30. Guidelines and Protocol Advisory Comittee, British Columbia Medical Services Commission. Genital tract cancers in females: ovarian, fallopian tube, and primary peritoneal cancers. 2014. https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/bc-guidelines/ovarian-cancers. Accessed 12 Jul 2018.
31. British Gynaecological Cancer Society. Epithelial ovarian / fallopian tube / primary peritoneal cancer guidelines: recommendations for practice. 2017. bgcs.org.uk/BGCS Guidelines Ovarian Guidelines 2017.pdf. Accessed 12 Jul 2018.
32. Ireland National Cancer Control Programme. Ovarian cancer GP referral for symptomatic women. 2016. www.hse.ie/eng/services/list/5/cancer/profinfo/resources/gpreferrals/Ovarian-Cancer-Referral-Guidelines.pdf. Accessed 29 Aug 2018.
33. National Institute for Health and Care Excellence. Suspected cancer: recognition and referral (NG12). 2015. www.nice.org.uk/guidance/ng12. Accessed 18 Jul 2019.
34. Healthcare Improvement Scotland. Scottish Referral Guidelines for Suspected Cancer. 2014. http://www.healthcareimprovementscotland.org/our_work/cancer_care_improvement/programme_resources/scottish_referral_guidelines.aspx. Accessed 12 Jul 2018.
35. Scottish Intercollegiate Guidelines Network. Management of epithelial ovarian cancer. 2013. http://www.sign.ac.uk/sign-135-management-of-epithelial-ovarian-cancer.html. Accessed 12 Jul 2018.
36. The Danish National Health Authority. Integrated ovarian cancer patient pathway. 2016. https://www.sst.dk/da/udgivelser/2016/~/media/765a142bc73446399c769a5d8b54c3dc.ash x. Accessed 12 Jul 2018.
37. The Norwegian Directorate of Health. Ovarian cancer patient pathway. 2016. https://helsedirektoratet.no/Retningslinjer/Pakkeforløp for eggstokkreft.pdf. Accessed 29 Aug 2018.
38. Maringe C, Walters S, Butler J, Coleman MP, Hacker N, Hanna L, et al. Stage at diagnosis and ovarian cancer survival: evidence from the international Cancer benchmarking partnership. Gynecol Oncol. 2012;127:75–82.
39. Rose PW, Rubin G, Perera-Salazar R, Almberg SS, Barisic A, Dawes M, et al. Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. BMJ Open. 2015;5:e007212.
40. Nicholson B, Mant D, Shinkins B, Rose P, Neal R, Hart N, et al. International variation in adherence to referral guidelines for suspected cancer: a secondary analysis of survey data. Br J Gen Pract. 2016;66:e106–13.
41. Giamougiannis P, Ebef NG. Dit, Mahmood T. the implications of new guideline 122 on the detection of ovarian cancer. Eur J Obstet Gynecol Reprod Biol. 2016;206:e154.
42. Neal RD, Din NU, Hamilton W, Ukoumunne OC, Carter B, Stapley S, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK general practice Research database. Br J Cancer. 2014;110:584–92.
43. Cancer Research UK. International Cancer Benchmarking Partnership, current research. https://www.cancerresearchuk.org/health-professional/data-and-statistics/international-cancer-benchmarking-partnership-icbp/current-research. Accessed 24 Aug 2018.
44. Maclean R, Jeffreys M, Ives A, Jones T, Verne J, Ben-Shlomo Y. Primary care characteristics and stage of cancer at diagnosis using data from the national cancer registration service, quality outcomes framework and general practice information. BMC Cancer. 2015;15:500.
45. Møller H, Gildea C, Meechan D, Rubin G, Round T, Vedsted P. Use of the English urgent referral pathway for suspected cancer and mortality in patients with cancer: cohort study. BMJ. 2015;351:h5102.
46. Rai N, Nevin J, Downey G, Abedin P, Balogun M, Kehoe S, et al. Outcomes following implementation of symptom triggered diagnostic testing for ovarian cancer. Eur J Obstet Gynecol Reprod Biol. 2015;187:64–9.
47. Baker LC, Chan J. Laws requiring health plans to provide direct access to obstetricians and gynecologists, and use of cancer screening by women. Health Serv Res. 2007;42:990–1007.
48. Green D, Irvine B. Health Care in France and Germany: lessons for the UK. Institute for the Study of Civil Society, London: Civitas; 2001.
49. Ebell MH, Culp MB, Radke TJ. A systematic review of symptoms for the diagnosis of ovarian Cancer. Am J Prev Med. 2016;50:384–94.
50. Goff BA, Mandel LS, Drescher CW, Urban N, Gough S, Schurman KM, et al. Development of an ovarian cancer symptom index: possibilities for earlier detection. Cancer. 2007;109:221–7.
51. Myers ER, Bastian LA, Havrilesky LJ, Kulasingam SL, Terplan MS, Cline KE, et al. Management of adnexal mass. Evid Rep Technol Assess (Full Rep). 2006;130:1–145.
52. Williams P, Bond C, Burton C, Murchie P. A systematic review of the use, quality and effects of pelvic examination in primary care for the detection of gynaecological cancer. J Obs Gynaecol. 2018;38:737.

53. Gilbert L, Basso O, Sampalis J, Karp I, Martins C, Feng J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. Lancet Oncol. 2012;13:285–91.
54. Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK collaborative trial of ovarian Cancer screening (UKCTOCS). Lancet Oncol. 2009;10:327–40.

Publisher's Note

PLOS MEDICINE

RESEARCH ARTICLE

The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study

Garth Funston[1]*, Willie Hamilton[2], Gary Abel[2], Emma J. Crosbie[3,4], Brian Rous[5], Fiona M. Walter[1]

1 The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, **2** University of Exeter, Exeter, United Kingdom, **3** Gynaecological Oncology Research Group, Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom, **4** Department of Obstetrics and Gynaecology, Manchester University NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, United Kingdom, **5** National Cancer Registration and Analysis Service, Public Health England, Cambridge, United Kingdom

* gf272@medschl.cam.ac.uk

OPEN ACCESS

Citation: Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM (2020) The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. PLoS Med 17(10): e1003295. https://doi.org/10.1371/journal.pmed.1003295

Academic Editor: Steven D. Shapiro, University of Pittsburgh, UNITED STATES

Received: February 3, 2020

Accepted: September 11, 2020

Published: October 28, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pmed.1003295

Data Availability Statement: Data cannot be shared publicly as they are analysed under licence. All data used in this study are available from CPRD

Abstract

Background

The serum biomarker cancer antigen 125 (CA125) is widely used as an investigation for possible ovarian cancer in symptomatic women presenting to primary care. However, its diagnostic performance in this setting is unknown. We evaluated the performance of CA125 in primary care for the detection of ovarian and non-ovarian cancers.

Methods and findings

We studied women in the United Kingdom Clinical Practice Research Datalink with a CA125 test performed between 1 May 2011–31 December 2014. Ovarian and non-ovarian cancers diagnosed in the year following CA125 testing were identified from the cancer registry. Women were categorized by age: <50 years and ≥50 years. Conventional measures of test diagnostic accuracy, including sensitivity, specificity, and positive predictive value, were calculated for the standard CA125 cut-off (≥35 U/ml). The probability of a woman having cancer at each CA125 level between 1–1,000 U/ml was estimated using logistic regression. Cancer probability was also estimated on the basis of CA125 level and age in years using logistic regression. We identified CA125 levels equating to a 3% estimated cancer probability: the "risk threshold" at which the UK National Institute for Health and Care Excellence advocates urgent specialist cancer investigation.

A total of 50,780 women underwent CA125 testing; 456 (0.9%) were diagnosed with ovarian cancer and 1,321 (2.6%) with non-ovarian cancer. Of women with a CA125 level ≥35 U/ml, 3.4% aged <50 years and 15.2% aged ≥50 years had ovarian cancer. Of women with a CA125 level ≥35 U/ml who were aged ≥50 years and who did not have ovarian cancer, 20.4% were diagnosed with a non-ovarian cancer. A CA125 value of 53 U/ml equated to a 3% probability of ovarian cancer overall. This varied by age, with a value of 104 U/ml in

(www.cprd.com). Permission to access data is through the Independent Scientific Advisory Committee (ISAC, contact via: isac@cprd.com). Derived data used to prepare article figures are available from the University of Cambridge Repository: www.repository.cam.ac.uk (https://doi.org/10.17863/CAM.56363).

Funding: This research arises from the CanTest Collaborative, which is funded by Cancer Research UK [C8640/A23385], of which GF is Clinical Research Fellow, GA is the Senior Statistician, and WH and FMW are Directors. The study was also funded by the National Institute of Health Research (NIHR) School of Primary Care Research [FR17 424] (GF, FMW). EJC is supported through the NIHR Manchester Biomedical Research Centre [IS-BRC-1215-20007]. The funders of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AUC, area under the curve; CA125, cancer antigen 125; CI, confidence interval; CT, computed tomography; GP, general practitioner; IT, information technology; NICE, National Institute for Health and Care Excellence; NPV, negative predictive value; PPV, positive predictive value; ROCkeTS, Refining Ovarian Cancer Test Accuracy Scores.

40-year-old women and 32 U/ml in 70-year-old women equating to a 3% probability. The main limitations of our study were that we were unable to determine why CA125 tests were performed and that our findings are based solely on UK primary care data, so caution is need in extrapolating them to other healthcare settings.

Conclusions

CA125 is a useful test for ovarian cancer detection in primary care, particularly in women ≥50 years old. Clinicians should also consider non-ovarian cancers in women with high CA125 levels, especially if ovarian cancer has been excluded, in order to prevent diagnostic delay. Our results enable clinicians and patients to determine the estimated probability of ovarian cancer and all cancers at any CA125 level and age, which can be used to guide individual decisions on the need for further investigation or referral.

Author summary

Why was this study done?

- CA125 is widely used as a test for ovarian cancer in women presenting with relevant symptoms, both in the UK and internationally.
- CA125 has been extensively evaluated in the specialist care setting and in screening studies but little was known about its diagnostic performance in primary care, where most patients with ovarian cancer first present.

What did the researchers do and find?

- We evaluated the diagnostic performance of CA125 in 50,780 women undergoing testing in English general practice.
- Of women with CA125 levels above the current abnormal cut-off, 10.1% were diagnosed with ovarian cancer and a further 12.3% with another form of cancer.
- Cancer was more common in women with abnormal CA125 levels if they were ≥50 years of age.
- We developed models to estimate the probability of ovarian cancer and all cancer based on a woman's age and CA125 level.

What do these findings mean?

- Clinicians should consider the possibility of non-ovarian cancer, in addition to ovarian cancer, in women with high CA125 levels.
- Our models will enable patients and clinicians to determine the estimated probability of cancer based on an individual's age and CA125 level.
- This information can be used to help make decisions about the need for further investigation or urgent referral to a specialist.

- The findings should also help inform guidelines, as they will allow recommendations for further testing to be made on the basis cancer probability rather than a single CA125 cutoff.

Introduction

Ovarian cancer is the eighth most common cancer to affect women worldwide, accounting for over 384,000 deaths in 2018 [1]. Survival depends on stage at diagnosis, with five-year net survivals of 93% for stage I, 68% for stage II, 27% for stage III, and 13.4% for stage IV disease [2]. Most women are diagnosed following a symptomatic presentation [3], and, in healthcare systems in which general practitioners (GPs) play a gatekeeping role, this initial presentation usually takes place in primary care [4].

Symptoms can occur at all stages of ovarian cancer [5]. However, they are usually nonspecific and are common in women without ovarian cancer, so they only have modest positive predictive values for the disease [5,6]. The serum biomarker cancer antigen 125 (CA125) is widely used in countries around the world, including the United States, Australia, Canada, and Ireland, as an investigation for ovarian cancer in symptomatic women presenting to primary care [7]. In 2011, the United Kingdom National Institute for Health and Care Excellence (NICE) recommended that women with symptoms of possible ovarian cancer be tested for CA125 in primary care, with further investigation advocated in those with CA125 levels $\geq$35 U/ml [8]. The chosen cutoff of 35 U/ml is the conventional upper limit of normal for CA125 and derives from a study in which 1% of healthy women and 82% of patients with ovarian cancer had a CA125 level >35 U/ml [9].

CA125 has been studied extensively in screening studies and in women in secondary care with pelvic masses but not in women presenting with symptoms of possible ovarian cancer in primary care. The NICE recommendations on CA125 testing for symptomatic women are based on extrapolated secondary care and screening data rather than primary care data [8]. The performance characteristics of a test vary with disease prevalence, disease severity, and the prevalence of other conditions that elevate test levels, so it is important to evaluate CA125 within the intended population [10].

When evaluating the diagnostic performance of a test such as CA125, it is standard practice to report accuracy characteristics, including the positive predictive value (PPV), after applying a particular cutoff. However, the PPV provides the "average" probability of disease for all women with a test level at or above the set cutoff rather than the probability of disease at a given test level. Knowledge of the probability of cancer at any given CA125 level is likely to be more clinically useful than the PPV, as it would allow patients and clinicians to interpret their individual CA125 test results, which could help guide decisions on the need for further investigations. NICE revised their cancer guidance in 2015, using a "risk threshold" of $\geq$3% as the threshold for urgent cancer investigation in symptomatic women, but ovarian cancer guidance, including the chosen CA125 cut-off of 35 U/ml, remained unchanged [11]. Knowledge of the estimated probability of cancer at each CA125 level could help inform health policy both in the UK and internationally.

The primary aim of this study was to explore the relationship between CA125 level and ovarian cancer probability, to identify the CA125 level at which a 3% probability of ovarian cancer was reached. Given the nonspecific nature of ovarian cancer symptoms, and reports

indicating CA125 is commonly elevated in other cancers [12,13], a second aim was to explore the relationship between CA125 level and the probability of all cancers. To allow comparison with existing literature on CA125 diagnostic accuracy, we also calculated conventional test performance characteristics, including PPV, sensitivity, and specificity, applying the standard cutoff (≥35U/ml).

Methods

Ethics statement

The study was approved by the Independent Scientific Advisory Committee (ISAC) for the Medicines and Healthcare Products Regulatory Agency (protocol number 18_184). All data were provided to researchers in an anonymized form, and individual consent was not required.

Data source

This was a retrospective cohort study using linked data from the Clinical Practice Research Datalink (CPRD) GOLD dataset and the National Cancer Registration and Analysis Service (NCRAS). The CPRD GOLD dataset contains anonymized, coded, primary care data including demographics, laboratory results, symptoms, and diagnoses for around 11 million patients. It is broadly representative of the UK population [14]. The NCRAS (English cancer registry) collects cancer registration data on patients, including detailed information on tumor topography, stage, and date of diagnosis. NCRAS obtains data from multiple sources including hospitals, GP surgeries, and death certificates and reports a near 100% case ascertainment [15]. Linkage of CPRD and NCRAS data was performed at a patient level by a third party, National Health Service (NHS) Digital [16]. As NCRAS only collects details of cancers diagnosed in England, the study was restricted to English general practices. The approved ISAC protocol, which covers several linked studies, is included in the S1 Text and S2 Text.

This report conforms to the STARD and RECORD statements [17,18]. A completed STARD checklist is included with this article (S3 Text).

Participants

We included women with a code for CA125 measurement in primary care (S1 Table) between 1 May 2011 and the 31 December 2014. There has never been a national ovarian cancer screening program in the UK, and the only indication for CA125 testing in English primary care is a presentation with a symptom of possible ovarian cancer. As such, we assumed that CA125-tested women were symptomatic.

Women who were <18 years old or registered at a GP practice not deemed "up-to-standard" on data quality by CPRD on the date of their first CA125 test during this period were excluded [14]. Women with a record of ovarian cancer in NCRAS data on or before the CA125 test date were also excluded, as were women with a CA125 test in the 12 months before the first CA125 test during the study period. Only CA125 entries recorded in standard equivalent units of CA125 measurement (U/ml, IU/ml, KU/L, KIU/L) were accepted. Although NICE recommends a CA125 cutoff of ≥35 U/ml, individual laboratory cutoffs varied. We excluded CA125 values associated with spurious laboratory cutoffs (245, 420, and 455 U/ml) and those where no cutoff was given. Subsequent sensitivity analyses, including CA125 entries recorded in all units and associated with all laboratory upper cutoffs, had minimal impact on our results. The first CA125 test during the study period was used in analyses.

Clinical outcomes

Primary outcome. Our primary clinical outcome was the diagnosis of ovarian cancer, as recorded using International Classification of Diseases (ICD)-10 codes in NCRAS data, in the 12 months following the initial CA125 test. With reference to the International Federation of Gynaecology and Obstetrics (FIGO) and WHO classifications [19,20], we defined ovarian cancer as an ovarian malignancy (C56), a fallopian tube malignancy (C57.0), a peritoneal malignancy (C48.1, C48.2), or a neoplasm of uncertain behavior of the ovary (D39.1). We assumed that cancer diagnosed within 12 months of the initial CA125 test was present at the time of testing. It is possible that incidental ovarian cancers may arise and be diagnosed in the year following testing or that it may take longer than 1 year from presentation in primary care to diagnosis. A period of 1 year, which has been used widely in similar studies [21,22], was chosen as a compromise between minimizing the inclusion of incidental cancers and maximizing the inclusion of relevant cancers.

Our primary outcome included borderline ovarian tumors, as these are treated collectively with invasive tumors in NICE recommendations on CA125 testing and generally require surgical management [23]. Although their timely detection in symptomatic women is important, borderline tumors are less likely to cause an elevation in serum CA125, and their prognosis is very good even if detected late [23]. We therefore performed a subanalysis in which invasive ovarian cancer formed the outcome.

Secondary outcome. Our secondary outcome was the diagnosis of non-ovarian cancers. The earliest record of cancer, excluding nonmelanoma skin cancers, was identified in the 12 months following initial CA125 testing in women without ovarian cancer. We refer to this group of cancers as "non-ovarian cancer." Where we discuss the combined non-ovarian and ovarian cancer groups, we use the term "all cancers."

Descriptive outcomes. In order to report the symptoms that may have triggered CA125 testing, symptoms coded in the 30 days before CA125 testing were identified from CPRD data using a code list of ovarian cancer symptoms from current NICE guidelines [11].

Ovarian cancer stage was determined using the Tumor Nodes Metastasis (TNM) staging system or, where not recorded, the FIGO staging system, and the proportions of ovarian cancers at each stage identified [19].

The histology of invasive ovarian tumors was identified from NCRAS data and categorized on the basis of ICD10 codes.

Statistical analysis

We calculated the PPV, negative predictive value (NPV), sensitivity, and specificity of CA125 for ovarian cancer at or above the current cut-off (35 U/ml). A nonparametric receiver operator characteristic (ROC) curve was constructed, and the area under the curve (AUC) determined. This analysis was repeated for invasive ovarian cancer and all cancers combined. After excluding ovarian cancer patients, it was also repeated for non-ovarian cancer. As ovarian cancer incidence is greater in older women and most cases occur in women post-menopause, we repeated all analyses for women $<$50 years and $\geq$50 years of age [24].

We used logistic regression to examine the relationship between CA125, as a continuous variable, and ovarian cancer diagnosis. CA125 level was highly skewed, and so it was log-transformed prior to regression analysis. Log CA125 was centered on a value of 3, the closest integer to the mean. The relationship between log CA125 level and ovarian cancer was nonlinear. To account for this, we used restricted cubic splines. As recommended by Harrell [25], we compared the Akaike Information Criterion (AIC) for models containing 3, 4, and 5 knots. The 5-knot model produced the smallest AIC and so was taken forward. Knots were placed at

standard, equally spaced percentiles of the marginal distribution of the variable (S4 Text) [25]. This regression model was used to predict the odds of cancer for a range of CA125 levels (1–1,000 U/ml), which were then converted into probabilities.

The logistic regression analysis was repeated for the <50 years and ≥50 years age groups. Significant differences were noted between these groups in terms of the estimated ovarian cancer probabilities. Given this, and on the recommendation of a peer reviewer, we constructed a multivariable regression model including age in years (mean centered) as a continuous variable and CA125 level, applying the same approach as described here previously. Five knots were included for each variable (S4 Text). This regression model was used to predict the odds of ovarian cancer for CA125 levels (1–1,000 U/ml) in women of different ages. Results for women aged 30, 40, 50, 60, 70, and 80 years of age are presented as examples in this paper.

All the aforementioned steps were repeated for our secondary outcome and for the invasive cancer subanalysis. Full details of all models are included in S4 Text.

Statistical software. All analyses were performed in Stata version 15.1 (StataCorp, www.stata.com). The DIAGT module was used to calculate summary diagnostic accuracy statistics [26]. All confidence intervals (CI) are reported at the 95% level.

Results

After exclusions, our cohort consisted of 50,780 women (**Fig 1**).

The ovarian cancer incidence in the cohort was 0.9% and was 3 times higher in the ≥50 years group than the <50 years group (**Table 1**). The median interval between CA125 testing and ovarian cancer diagnosis was 42 days (interquartile range: 25–62 days) and the mean patient age was 56 years (range: 18–102 years).

Cancer stage

Of the 456 ovarian cancers, 172 (37.7%) were stage I or II, and 209 (45.8%) were stage III or IV. No stage was recorded in 75 (16.4%) cases (S2 Table).

Cancer morphology and histology

Of the ovarian cancers diagnosed, 21.5% ($n = 98$) were borderline tumors. The proportion of malignancies that were borderline varied with age, with 50% of tumors in the <50 years group and 15.4% in the ≥50 years group being borderline (S3 Table). Serous epithelial tumors were the most common tumor type, accounting for 48.6% ($n = 174$) of invasive tumors. In the <50 years group, 12.5% ($n = 5$) of invasive tumors were of nonepithelial origin compared with 2.5% ($n = 8$) in the ≥50 years group.

Recorded symptoms

Symptoms of possible ovarian cancer were recorded for 24,269 women (47.8%) on the same day or in the 30 days preceding CA125 testing; the most common was abdominal pain (**Table 2**). Multiple symptoms were recorded in 1,477 (6.1%) women.

Diagnostic performance applying the standard cutoff (≥35 U/ml)

The diagnostic performance characteristics of CA125 were calculated after applying the standard cutoff (≥35 U/ml) (**Table 3**). At or above the 35 U/ml cutoff, CA125 demonstrated a PPV of 10.1% (95% CI 9.1–11.2), an NPV of 99.8% (95% CI 99.7–99.8), a sensitivity of 77.0% (95% CI 72.8–80.8%) and a specificity of 93.8% (95% CI 93.6–94.0) for ovarian cancer. The AUC was 0.92 (95% CI 0.90–0.93). The AUC was greater in the ≥50 years group (AUC: 0.93,

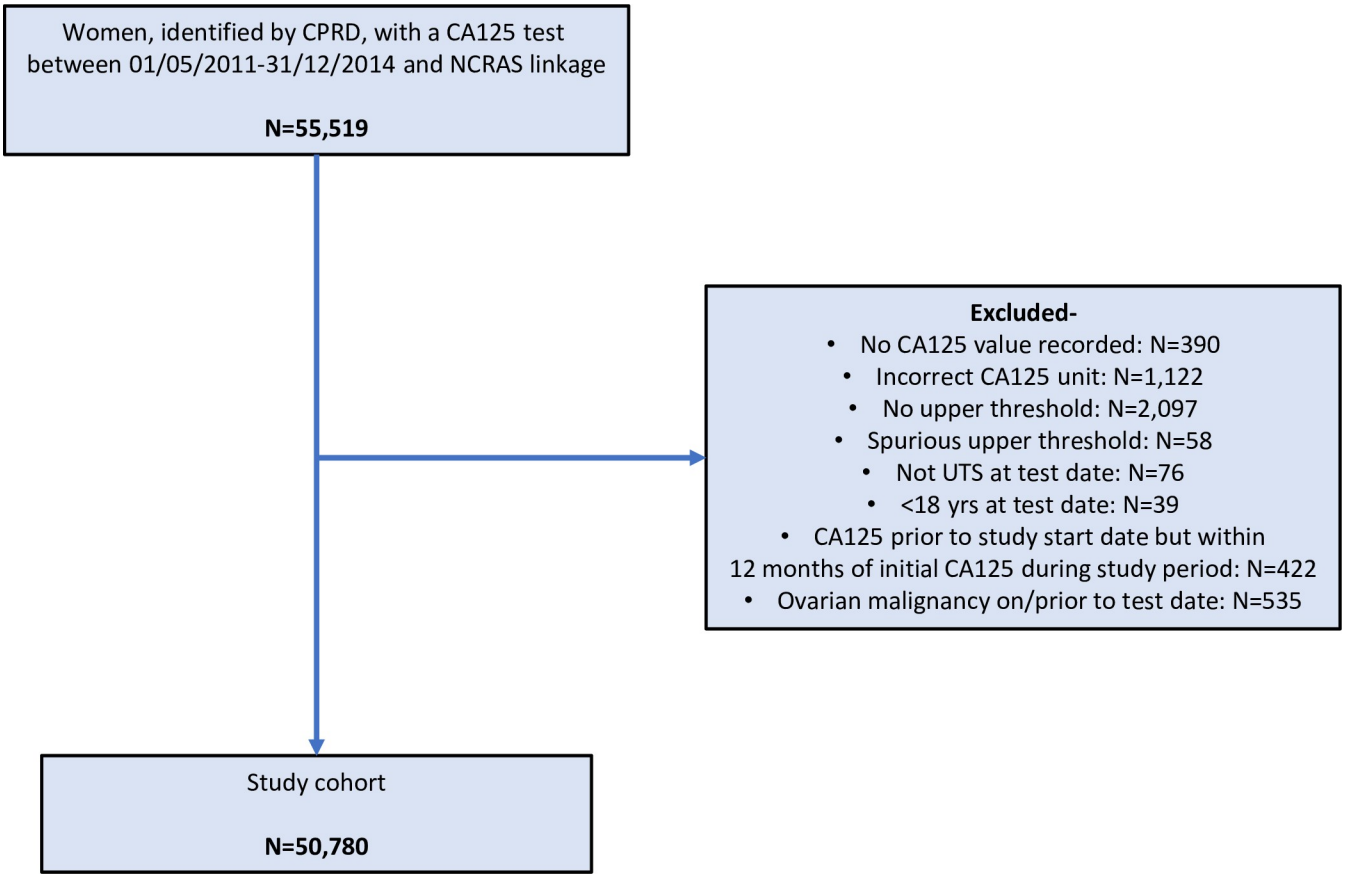


Fig 1. Flow diagram illustrating the identification of the study cohort and application of exclusion criteria. UTS is a quality metric, provided by CPRD, which indicates if the data from a GP practice are of sufficient quality to be used in research [14]. CA125, cancer antigen 125; CPRD, Clinical Practice Research Datalink; NCRAS, National Cancer Registration and Analysis Service; UTS, up to standard.

https://doi.org/10.1371/journal.pmed.1003295.g001

95% CI 0.92–0.95) than the <50 years group (AUC 0.86, 95% CI 0.82–0.91) and the PPV, sensitivity and specificity were also higher in the ≥50 group.

When the outcome was restricted to invasive ovarian cancers, CA125 demonstrated a slightly lower PPV (8.8%, 95% CI 7.8–9.8), and a higher sensitivity (84.9%, 95% CI 80.8–88.5) (**Table 3**).

Of the 50,324 women without ovarian cancer, 1,321 (2.6%) were diagnosed with a non-ovarian cancer. The incidence of non-ovarian cancers in women with a CA125 <35 U/ml was

Table 1. Patient numbers, incidence of raised CA125 tests (≥35 U/ml) and cancer incidence by age group.

	<50 years	≥50 years	Overall cohort
Number of patients, *N*	19,694	31,086	50,780
Raised (≥35 U/ml) CA125, *N* (%)	1,482 (7.5)	1,986 (6.4)	3,468 (6.8)
Ovarian cancers, *N* (%)	80 (0.4)	376 (1.2)	456 (0.9)
Non-ovarian cancer, *N* (%)	161 (0.8)	1,160 (3.7)	1,321 (2.6)

CA125, cancer antigen 125.

https://doi.org/10.1371/journal.pmed.1003295.t001

Table 2. Ovarian cancer symptoms and signs coded in the 30 days prior to CA125 testing.

Symptom/sign	Patients, *N* (%)
Abdominal pain	11,933 (49.2)
Abdominal distension or bloating	5,686 (23.4)
Change in bowel habit	2,866 (11.8)
Fatigue	1,692 (7.0)
Pelvic pain	1,632 (6.7)
Weight loss	913 (3.8)
Urinary frequency	552 (2.3)
Abdominal or pelvic mass	419 (1.7)
Loss of appetite	113 (0.5)
Urinary urgency	86 (0.4)
Ascites	26 (0.1)

% is the proportion of patients with a given symptom out of the total number of patients who have a coded symptom (*N* = 24,269). Categories are not mutually exclusive: a patient may have had more than 1 symptom coded.
CA125, cancer antigen 125.

https://doi.org/10.1371/journal.pmed.1003295.t002

Table 3. Performance characteristics of CA125 for ovarian cancer, invasive ovarian cancer, non-ovarian cancers and all cancers.

Cancer	Group	PPV, % (95% CI)	NPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC (95% CI)
Ovarian	All ages	10.1 (9.1–11.2)	99.8 (99.7–99.8)	77.0 (72.8–80.8)	93.8 (93.6–94.0)	0.92 (0.90–0.93)
	<50 years	3.4 (2.5–4.4)	99.8 (99.8–99.9)	62.5 (51.0–73.1)	92.7 (92.3–93.1)	0.86 (0.82–0.91)
	≥50 years	15.2 (13.6–16.8)	99.7 (99.7–99.8)	80.1 (75.7–84.0)	94.5 (94.3–94.8)	0.93 (0.92–0.95)
Ovarian: invasive	All ages	8.8 (7.8–9.8)	99.9 (99.9–99.9)	84.9 (80.8–88.5)	93.7 (93.5–93.9)	0.94 (0.92–0.96)
	<50 years	2.0 (1.3–2.8)	99.9 (99.9–100)	72.5 (56.1–85.4)	92.6 (92.2–93.0)	0.88 (0.82–0.95)
	≥50 years	13.8 (12.4–15.4)	99.9 (99.8–99.9)	86.5 (82.2–90.0)	94.4 (94.2–94.7)	0.95 (0.93–0.97)
Non-ovarian	All ages	12.3 (11.2–13.5)	98.0 (97.9–98.1)	29.1 (26.6–31.6)	94.4 (94.2–94.6)	0.68 (0.66–0.69)
	<50 years	2.8 (2.0–3.8)	99.3 (99.2–99.4)	24.8 (18.4–32.3)	92.8 (92.5–93.2)	0.62 (0.58–0.67)
	≥50 years	20.4 (18.5–22.4)	97.2 (97.0–97.4)	29.7 (27.0–32.4)	95.5 (95.2–95.7)	0.70 (0.69–0.72)
All cancers	All ages	21.2 (19.8–22.6)	97.8 (97.7–97.9)	41.4 (39.1–43.7)	94.4 (94.2–94.6)	0.74 (0.73–0.75)
	<50 years	6.1 (4.9–7.4)	99.2 (99.0–99.3)	37.3 (31.2–43.8)	92.8 (92.5–93.2)	0.70 (0.67–0.74)
	≥50 years	32.5 (30.4–34.6)	96.9 (96.7–97.1)	42.0 (39.5–44.5)	95.5 (95.2–95.7)	0.76 (0.75–0.78)

PPV, NPV, sensitivity and specificity are calculated for a cutoff of ≥35 U/ml. Accuracy characteristics for "non-ovarian" cancer were calculated following exclusion of patients with ovarian cancer.
AUC, area under the curve; CA125, cancer antigen 125; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

https://doi.org/10.1371/journal.pmed.1003295.t003

Table 4. Cancers diagnosed in women without ovarian cancer.

Cancer type (ICD10 codes)	*N*	*N* < 35 U/ml (%)	*N* ≥ 35 U/ml (%)
Unknown primary (C80)	46	8 (17)	38 (83)
Secondary: respiratory and digestive (C78)	23	9 (39)	14 (61)
Pancreas (C25)	93	47 (51)	46 (49)
Lung (C34)	104	55 (53)	49 (47)
Liver, biliary (C22, C23, C24)	34	21 (62)	13 (38)
Uterus (C54, C55, D39.0)	132	84 (64)	48 (36)
Upper GI (C15, C16, C17, D37.1, D37.2)	66	46 (70)	20 (30)
Lower GI (C18, C19, C20, C21,D37.3,D37.4,D37.5)	255	197 (77)	58 (23)
Hematological (C81, C82, C83, C84, C85, C90, C91, C92, C96, D45, D46, D47)	112	83 (74)	29 (26)
Kidneys, urinary tract (C64, C65, C66, C67, D41)	78	65 (83)	13 (17)
Breast (C50)	154	142 (92)	12 (8)
Other	224	180 (80)	44 (20)
Total	**1,321**	**937 (71)**	**384 (29)**

"Other" consists of cancers with fewer than 10 cases with CA125 values ≥35 U/ml. The cancers included in this group and their frequencies are shown in S4 Table.

CA125, cancer antigen 125; ICD-10, International Classification of Diseases, 10th revision.

https://doi.org/10.1371/journal.pmed.1003295.t004

2.0%, whereas the incidence in women with a CA125 ≥35 U/ml, which equates to the PPV for non-ovarian cancers, was 12.3% (95% CI 11.2–13.5) (**Table 3**). This varied markedly between the <50 years group (PPV 2.8%, 95% CI 2.0–3.8) and ≥50 years group (PPV 20.4%, 95% CI 18.5–22.4). The PPV for all cancers was 21.2% (95% CI 19.8–22.6). Almost half of patients diagnosed with pancreatic and lung cancer in our cohort had CA125 levels ≥35 U/ml (**Table 4**).

The probability of cancer by CA125 level

Fig 2 shows the relationship between CA125 level and the estimated probability of cancer, derived from logistic regression analyses. A CA125 level of 53 U/ml equated to a probability of 3% (95% CI 2.6–3.5) for ovarian cancer, whereas a CA125 level of 18 U/ml equated to a probability of 3% (95% CI 2.8–3.2) for all cancer. In a subanalysis in which invasive ovarian cancer formed the outcome, a CA125 level of 68 U/ml equated to a 3% probability.

Repeating the analysis in the <50 years and ≥50 years groups revealed that a much higher CA125 level was required to reach the 3% probability for ovarian cancer in the <50 years group (89 U/ml) than ≥50 years group (39 U/ml) (S4 Fig).

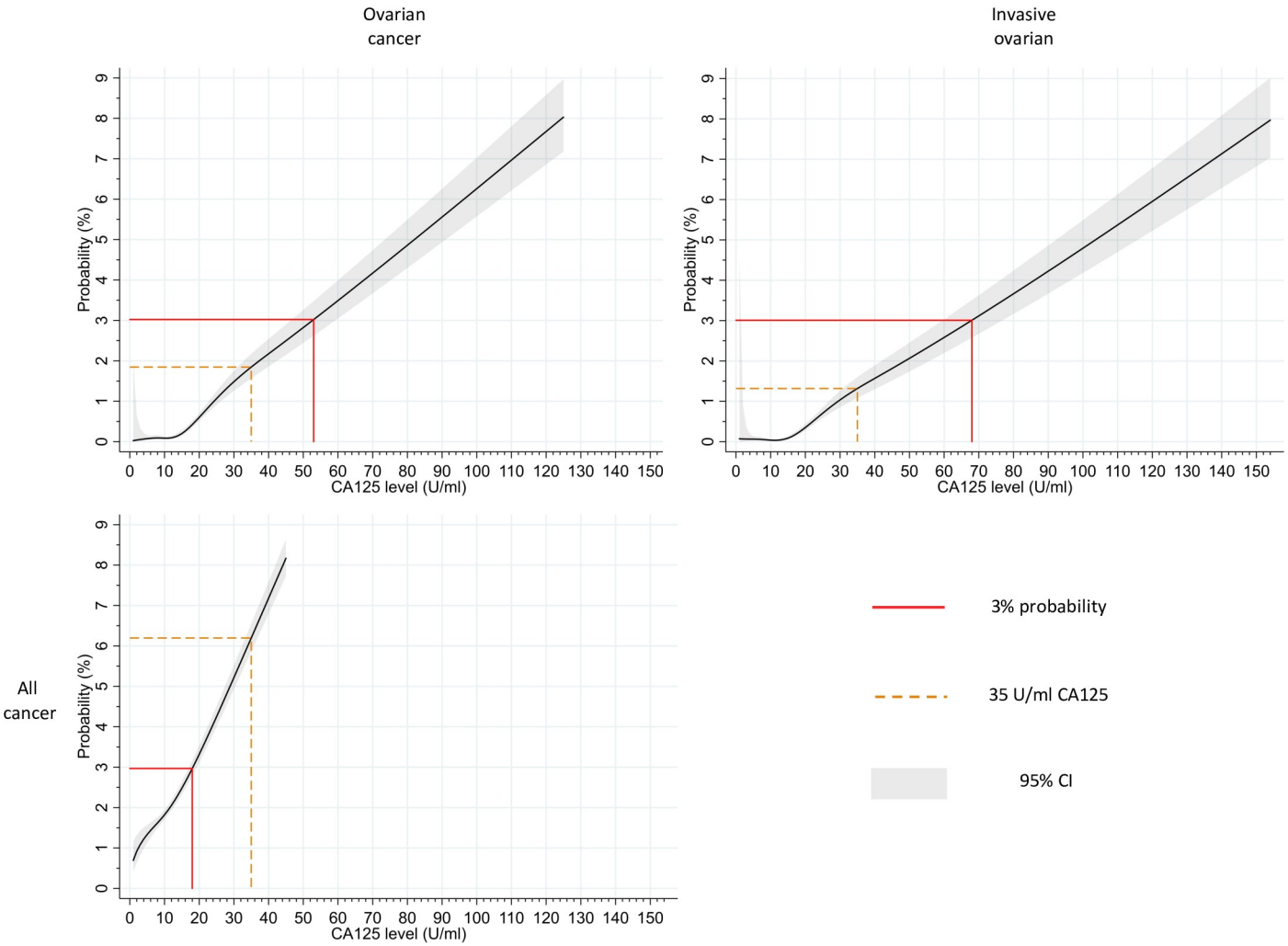


Fig 2. Relationship between CA125 level and estimated probability of ovarian cancer, invasive ovarian cancer, and all cancers. Estimated probabilities up to 8% are shown for ovarian cancer, invasive ovarian cancer, and all cancers. CA125 levels that correspond to the closest integer probabilities of 3% are indicated. The probabilities, which equate to a CA125 level of 35 U/ml, are also marked. Confidence intervals (95%) are displayed. Graphs showing probabilities at an extended range of CA125 values (up to 500 U/ml) for ovarian cancer, invasive ovarian cancer, and all cancer are included in S1 Fig, S2 Fig, and S3 Fig, respectively. Data used to construct these graphs (up to a CA125 level of 1,000 U/ml) are available via the University of Cambridge Repository [27]. CA125, cancer antigen 125; CI, confidence interval.

https://doi.org/10.1371/journal.pmed.1003295.g002

The probability of ovarian cancer by age and CA125 level

Fig 3 illustrates the relationship between CA125 level and the estimated probability of ovarian cancer at specific ages, derived from a logistic regression analysis. The probability of ovarian cancer at a given CA125 level varied markedly by age. The CA125 level required to reach the 3% ovarian cancer probability threshold fell from 104 U/ml in 40-year-old women to 32 U/ml in 70-year-old women. Similar age trends were noted when the analysis was repeated for invasive ovarian cancer (S5 Fig) and all cancer (S6 Fig).

Discussion

In this cohort study of over 50,000 women who underwent CA125 testing in English general practice, 10.1% of those with a CA125 at or above the conventional cutoff (35 U/ml) were

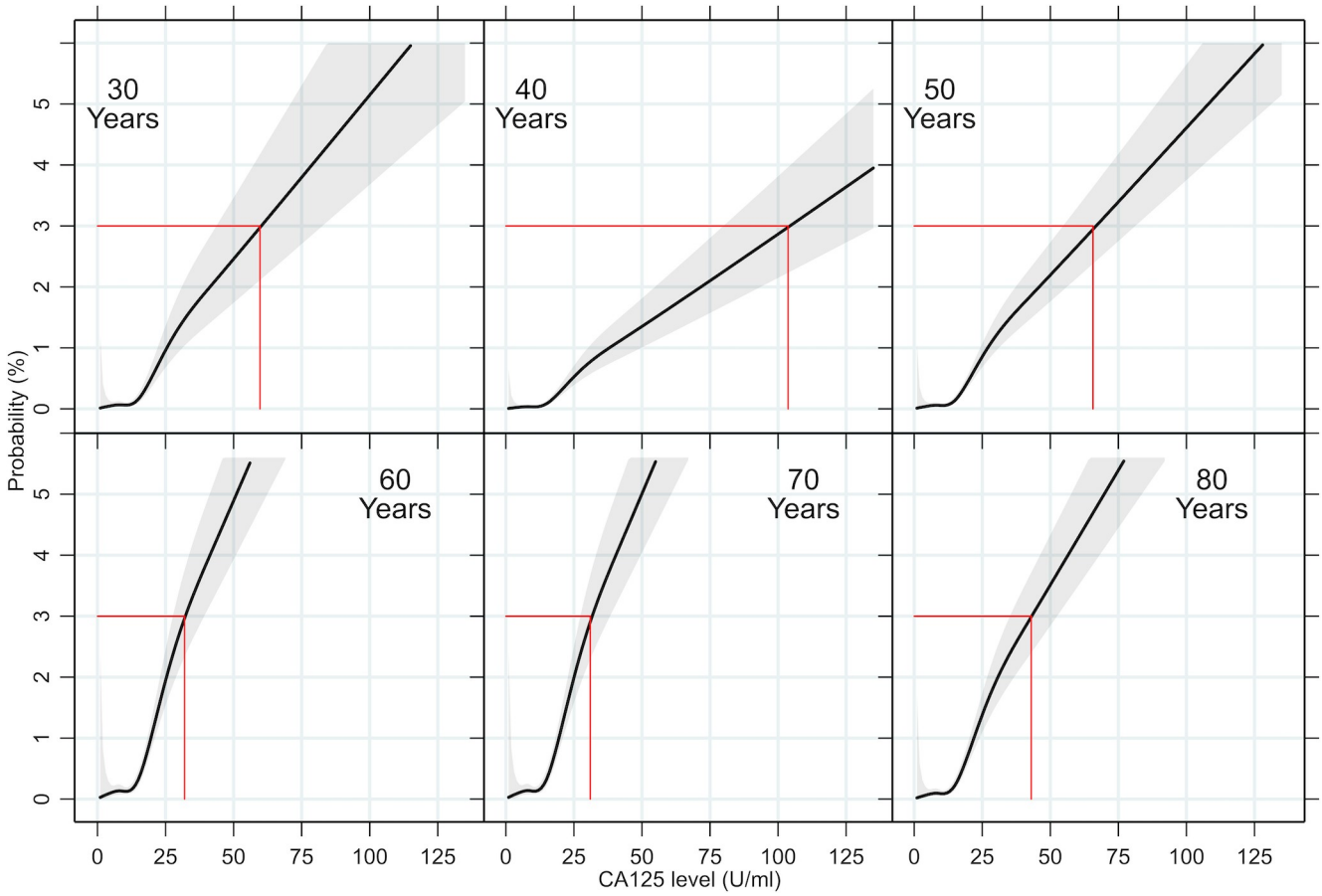


Fig 3. Relationship between CA125 level and estimated probability of ovarian cancer for women of different ages. Estimated ovarian cancer probabilities are shown in relation to CA125 level for women of 30, 40, 50, 60, 70, and 80 years of age. CA125 levels that correspond to the closest integer probabilities of 3% are indicated in red. Confidence intervals (95%) are displayed. Data used to construct these graphs (up to a CA125 level of 1,000 U/ml) are available via the University of Cambridge Repository [27]. CA125, cancer antigen 125.

https://doi.org/10.1371/journal.pmed.1003295.g003

diagnosed with ovarian cancer, and 12.3% were diagnosed with a different cancer. Almost a third of women aged ≥50 years with a CA125 ≥35 U/ml were diagnosed with some form of cancer. A CA125 level of 53 U/ml equated to an overall ovarian cancer probability of 3%—the threshold at which the UK NICE advocates urgent investigation or referral in symptomatic women. Marked variation was noted between women of different ages, with the 3% probability reached at lower CA125 levels in 70-year-old women than younger or older women.

Study limitations

This study relied on coded routinely collected data, so it was not possible to determine exactly why CA125 tests were requested. However, the only indication for CA125 testing in UK primary care is to investigate symptoms of possible ovarian cancer. We identified symptoms recorded before CA125 testing, which may have been the trigger for testing. In contrast to CA125 results, which are automatically transferred into the GP system from laboratories, symptoms are not always coded but instead are often recorded in the free text within the GP record, which cannot be accessed for research purposes [28]. We did not restrict our analysis to women with a coded ovarian cancer symptom as this could introduce bias, given that symptoms are more likely to be coded if they are severe or persistent [28].

Our results reflect real-world use of CA125 in English general practice. How CA125 is used in primary care in other countries may differ from practice in England. Baseline CA125 levels may also be affected by population characteristics, such as ethnicity, smoking status, and past medical history [29]. We did not include these variables in our analysis as our aim was to develop simple models that allow the estimated probability of cancer to be reported alongside the CA125 result in general practice, without the need to collect further detailed information from the patient. Although this study has significant international relevance, caution is needed when translating our findings to other countries and healthcare systems.

We report ovarian cancer stage at diagnosis, but CA125 diagnostic accuracy was not analyzed by stage. As CA125 tests were performed at variable intervals in the 12 months preceding diagnosis, such an analysis is likely to be misleading as an unknown number of cancers will have progressed during that period.

We have employed restricted cubic splines to model the nonlinear nature of the relationship between both age and CA125 level and cancer diagnoses. Although these provide a flexible approach to parameterizing the fitted relationships, there is a large degree of uncertainty at the extremes of age and CA125 level, and so the cancer probabilities for very old and young women and those with very low CA125 levels should be treated with caution, and the large CI noted.

Results in the context of other studies

In their 2011 ovarian cancer guidelines, NICE estimated that 0.81% of symptomatic primary care women with a CA125 $\geq$35 U/ml would have ovarian cancer [8]. Economic modeling and the recommendation for sequential testing with CA125 followed by ultrasound if the CA125 were abnormal was predicated on this estimate. Our findings indicate that the PPV is more than 12 times higher than estimated. This is consistent with the only other UK report of the PPV of CA125 in primary care, which found that 16 out of 152 women (11%) with a raised CA125 level had ovarian cancer [13]. The sensitivity of CA125 for ovarian cancer in our study was slightly lower—and the specificity higher—than reported in studies in which testing was performed in women with a pelvic mass prior to surgery in secondary care [30]. This is to be expected as tests generally have lower sensitivity and higher specificity in populations with a lower disease prevalence—the spectrum effect [10]. As anticipated, the PPVs for ovarian cancer in our cohort were lower than in secondary care patients with pelvic masses [31] and higher than in asymptomatic screening populations [32].

One of the most striking findings in our study was the high incidence of non-ovarian cancers in those with elevated CA125 levels, particularly in women aged 50 years or older. This reflects the nonspecific nature of ovarian cancer symptoms and also that CA125 is frequently raised in women with a variety of non-ovarian malignancies [12]. Crawford and colleagues reported that 16 out of 152 women (11%) referred from primary care with a raised CA125 were diagnosed with a non-ovarian cancer [13]. Furthermore, in asymptomatic screening populations, a higher incidence of non-ovarian cancers has been noted in women with raised CA125 levels (6.9%) than with normal CA125 levels (1.6%) [33].

We found that the estimated probability of ovarian cancer for a given CA125 level rose with age to peak in women in their seventies, which mirrors UK age-specific cancer incidence rates [34]. The exception was very young women—the probability of ovarian cancer at a given CA125 level was higher in women aged 30 than aged 40. This probably reflects GP testing practices in very young women (in whom ovarian cancer is extremely rare), with GPs having a strong reason to request a CA125 test in these women, thereby raising the pretest probability.

Clinical interpretation of the findings

Of the CA125 tests performed, 39% were in women <50 years of age; however, ovarian cancer is predominantly a disease of older and postmenopausal women. This is reflected in our findings, as only 18% of ovarian cancers and 11% of the invasive subtype occurred in women under 50. All measures of test performance, save for the NPV, were worse in women under 50 years than 50 years and over, even when borderline malignancies (which were more common in the younger age group) were excluded. A greater proportion of invasive tumors in the <50 years group were mucinous epithelial and nonepithelial cancers, both of which have less propensity to elevate serum CA125 than other ovarian cancer types, likely contributing to poorer test performance in the younger age group [31]. The results of our regression analysis indicate that, overall, only 1 in 110 women <50 years with a CA125 of exactly 35 U/ml will have an ovarian cancer, and only 1 in 308 will have an invasive subtype. Investigating younger women for ovarian cancer when there is high suspicion is important, but given the low incidence of ovarian cancer and relatively poor test performance in women under 50 years, CA125 tests should be performed and interpreted with caution in this group.

The total number of non-ovarian cancers diagnosed in women with raised CA125 levels exceeded that of ovarian cancers, but the numbers of women with each type of non-ovarian cancer was small. In isolation, CA125 is unlikely to be a useful test for the detection of individual types of non-ovarian cancer in primary care, most of which have superior triage tests. However, given our study findings, a high CA125 level in a woman ≥50 years should raise a suspicion of non-ovarian cancer. Clinicians should consider these cancers and whether further investigation is required, particularly if ovarian cancer has been excluded. Research is needed to determine the most appropriate follow-up and testing strategy for these women in order to ensure prompt diagnosis.

When assessing test performance, it is standard practice to evaluate test characteristics using a cutoff, above which the test is deemed abnormal and below which it is deemed normal. As per convention, we have presented this for CA125, applying the standard ≥35 U/ml cutoff. However, where the probability of having a disease varies markedly with the test level, PPV is of limited value in informing decisions about individual patients, as it effectively provides an average probability of disease for all women with "abnormal" results. In this study, women with very high CA125 values had a very high probability of being diagnosed with cancer. Conversely, those with CA125 levels around the 35 U/ml cutoff had a much lower probability of being diagnosed with cancer than the PPV would appear to indicate. In this study, we have quantified the risk of cancer in individuals with specific CA125 values at specific ages. This should be of much more use clinically than PPVs.

Estimated cancer probabilities will allow women and clinicians to interpret their individual CA125 result and could inform health policy both in the UK and internationally. For example, NICE currently recommend that women with a CA125 ≥35 U/ml, whether 35 U/ml or 1,000 U/ml, should be referred for an ultrasound scan, whereas no further investigations for ovarian cancer are advocated in women with levels below the cutoff. Instead, our results could be used to triage women of different ages, selecting those with a high probability of ovarian cancer for expedited referral and investigation. Women with a probability in excess of the NICE risk threshold of 3% could be referred via the urgent cancer pathway for specialist gynecology assessment and/or imaging. Women with lower probabilities might, after discussion between clinician and patient, be investigated using routine ultrasound, recognizing the fact that patients would opt for cancer testing at risk levels as low as 1% [35]. As only a woman's age and CA125 level are required to determine the cancer probability from our results, this information could readily be incorporated into laboratory information technology (IT) systems,

reported alongside the CA125 level, and communicated to patients in clear terms, e.g., "1 in 30 women of your age who have the same CA125 level in general practice will have ovarian cancer."

Although we have focused on the UK NICE 3% probability threshold for urgent cancer referral, our results would also allow alternative thresholds for referral to be implemented. A lower probability threshold may lead to the detection of more cancers, but this would come at the cost of larger numbers of cancer-free women being referred and further investigated, which can have negative consequences such as increased patient anxiety and financial cost [36]. Conversely, employing a higher probability threshold would lead to fewer cancer-free women being referred unnecessarily, but more cancers may be missed. A full health economic evaluation would greatly improve understanding of the implications of applying different referral thresholds.

The Refining Ovarian Cancer Test Accuracy Scores (ROCkeTS) study, a large ongoing prospective study in the UK evaluating a range of diagnostic tests and algorithms for ovarian cancer in secondary care, may provide insight into the most appropriate post-CA125 testing strategy [37]. Any such strategy should take account of the high incidence of non-ovarian cancers in women with high CA125 levels, as pelvic ultrasound alone will miss many of these malignancies. Other imaging modalities such as computed tomography (CT), which can detect multiple CA125 elevating cancers including ovarian, lung, and pancreatic cancer [38–40] and which is already used in several countries to investigate symptomatic women with elevated CA125 levels [7], could be appropriate. Further research is also needed to determine whether CA125 re-testing in primary care should be performed in women who have a normal ultrasound scan but persistent symptoms, as there is evidence from screening studies that a rising CA125 is associated with a higher risk of ovarian cancer, even if ultrasound is normal [41].

Conclusions

CA125 is a useful test for detecting ovarian cancer in primary care, particularly in women aged 50 years and over. Given the high incidence of non-ovarian cancers in women with elevated CA125 levels, clinicians should consider alternative cancers particularly when ovarian cancer has been excluded. The results of this study will enable patients and clinicians to interpret their CA125 result in terms of the probability of cancer at the pertinent CA125 level and age. The findings will also allow policy makers to provide recommendations for post-CA125 investigations on the basis of the probability of undiagnosed cancer, which could enable the expedited investigation and referral of those women most likely to have a cancer.

Supporting information

S1 Text. ISAC protocol. ISAC, Independent Scientific Advisory Committee.
(PDF)

S2 Text. Minor amendment to ISAC protocol (Dated: 02/07/2019). ISAC, Independent Scientific Advisory Committee.
(PDF)

S3 Text. Completed STARD checklist. STARD, Standards for Reporting of Diagnostic Accuracy Studies.
(PDF)

S4 Text. Logistic regression model specifications.
(PDF)

S1 Table. Read codes and terms used to identify CA125-tested women.
(PDF)

S2 Table. Ovarian cancer by stage of diagnosis.
(PDF)

S3 Table. Behavior and histology of ovarian tumors by age group (<50 years and ≥50 years).
(PDF)

S4 Table. Frequencies of non-ovarian cancers included in the "other" group in Table 4.
(PDF)

S1 Fig. Estimated probabilities of ovarian cancer at an extended range of CA125 levels.
(PDF)

S2 Fig. Estimated probabilities of invasive ovarian cancer at an extended range of CA125 levels.
(PDF)

S3 Fig. Estimated probabilities of all cancer at an extended range of CA125 levels.
(PDF)

S4 Fig. Relationship between CA125 level and estimated probability of ovarian cancer, invasive ovarian cancer and all cancers in women <50 years and ≥50 years of age. Estimated probabilities up to 8% for each cancer type in women <50 and ≥50 years of age are shown, save for invasive cancer in the <50 years group, in which a CA125 of 317 U/ml was required to reach an 8% probability.
(PDF)

S5 Fig. Relationship between CA125 level and estimated probability of invasive ovarian cancer for women of different ages. Probabilities are shown in relation to CA125 level for women of 30, 40, 50, 60, 70, and 80 years of age. CA125 levels that correspond to the closest integer probabilities of 3% are indicated in red (not displayed in 40 years of age- reached at 191 U/ml). The 95% confidence intervals are displayed.
(PDF)

S6 Fig. Relationship between CA125 level and estimated probability of all cancer for women of different ages. Probabilities are shown in relation to CA125 level for women of 30, 40, 50, 60, 70, and 80 years of age. CA125 levels that correspond to the closest integer probabilities of 3% are indicated in red. The 95% confidence intervals are displayed.
(PDF)

Acknowledgments

This work uses data provided by patients and collected by the NHS as part of their care and support. We are grateful to Mrs. Margaret Johnson for her input into study development and to Professor Greg Rubin for his helpful comments during paper preparation.

The views expressed are those of the authors and not necessarily those of the National Institute for Health Research (NIHR), the Department of Health and Social Care, or Cancer Research UK.

Author Contributions

Conceptualization: Garth Funston, Willie Hamilton, Emma J. Crosbie, Fiona M. Walter.

Data curation: Garth Funston, Brian Rous.

Formal analysis: Garth Funston, Gary Abel.

Funding acquisition: Garth Funston, Willie Hamilton, Fiona M. Walter.

Investigation: Garth Funston.

Methodology: Garth Funston, Willie Hamilton, Gary Abel, Brian Rous, Fiona M. Walter.

Project administration: Garth Funston.

Supervision: Fiona M. Walter.

Writing – original draft: Garth Funston.

Writing – review & editing: Garth Funston, Willie Hamilton, Gary Abel, Emma J. Crosbie, Brian Rous, Fiona M. Walter.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68:394–424. https://doi.org/10.3322/caac.21492 PMID: 30207593
2. Cancer Research UK. Ovarian cancer survival statistics. [cited 20 May 2020]. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three
3. Nagle CM, Francis JE, Nelson AE, Zorbas H, Luxford K, de Fazio A, et al. Reducing Time to Diagnosis Does Not Improve Outcomes for Women With Symptomatic Ovarian Cancer: A Report From the Australian Ovarian Cancer Study Group. J Clin Oncol. 2011; 29:2253–2258. https://doi.org/10.1200/JCO.2010.32.2164 PMID: 21537035
4. National Cancer Intelligence Network. Routes to Diagnosis 2006–2016 by year, V2.1a. 2019 [cited 2020 May 21]. Available from: http://www.ncin.org.uk/publications/routes_to_diagnosis
5. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. BMJ. 2009; 339: b2998. https://doi.org/10.1136/bmj.b2998 PMID: 19706933
6. Sundar S, Neal R, Kehoe E. Diagnosis of ovarian cancer. BMJ. 2015; 351:h5543. https://doi.org/10.1136/bmj.h5543 PMID: 26489957
7. Funston G, Melle VM, Ladegaard Baun M-L, Jensen H, Helpser C, Emery J, et al. Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. BMC Cancer. 2019; 19:1028. https://doi.org/10.1186/s12885-019-6211-2 PMID: 31676000
8. National Collaborating Centre for Cancer. Ovarian Cancer: The recognition and initial management of ovarian cancer (NICE Clinical Guidelines No. 122). Cardiff (UK); 2011.
9. Bast RC, Klug TL, John E St., Jenison E, Niloff JM, Lazarus H, et al. A Radioimmunoassay Using a Monoclonal Antibody to Monitor the Course of Epithelial Ovarian Cancer. N Engl J Med. 1983; 309:883–887. https://doi.org/10.1056/NEJM198310133091503 PMID: 6310399
10. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ. 2016; 353:i3139. https://doi.org/10.1136/bmj.i3139 PMID: 27334281
11. National Institute for Health and Care Excellence. Suspected cancer: recognition and referral (NG12). 2015 [cited 2019 Jul 18]. Available from: www.nice.org.uk/guidance/ng12
12. Duffy MJ, Bonfrer JM, Kulpa J, Rustin GJS, Soletormos G, Torre GC, et al. CA125 in ovarian cancer: European Group on Tumor Markers guidelines for clinical use. Int J Gynecol Cancer. 2005; 15:679–691. https://doi.org/10.1111/j.1525-1438.2005.00130.x PMID: 16174214
13. Crawford SM, Evans C. Outcome of elevated CA125 values from primary care following implementation of ovarian cancer guidelines. Fam Pract. 2017; 28:28.
14. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015; 44:827–836. https://doi.org/10.1093/ije/dyv098 PMID: 26050254

15. Public Health England. National Cancer Registration and Analysis Service (NCRAS). 2016 [cited 18 Jul 2019]. Available from: https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras

16. Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. Eur J Epidemiol. 2019; 34:91–99. https://doi.org/10.1007/s10654-018-0442-4 PMID: 30219957

17. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. Clin Chem. 2015; 351:h5527.

18. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015; 12:e1001885–e1001885. https://doi.org/10.1371/journal.pmed.1001885 PMID: 26440803

19. Prat J. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. Int J Gynaecol Obstet. 2014; 124:1–5. https://doi.org/10.1016/j.ijgo.2013.10.001 PMID: 24219974

20. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO Classification of Tumours of Female Reproductive Organs. Fourth Edition. Lyon: International Agency for Research on Cancer; 2014.

21. Watson J, Salisbury C, Banks J, Whiting P, Hamilton W. Predictive value of inflammatory markers for cancer diagnosis in primary care: a prospective cohort study using electronic health records. Br J Cancer. 2019; 120:1045–1051. https://doi.org/10.1038/s41416-019-0458-x PMID: 31015558

22. Neal RD, Din NU, Hamilton W, Ukoumunne OC, Carter B, Stapley S, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. Br J Cancer. 2014; 110:584–92. https://doi.org/10.1038/bjc.2013.791 PMID: 24366304

23. Fischerova D, Zikan M, Dundr P, Cibula D. Diagnosis, treatment, and follow-up of borderline ovarian tumors. Oncologist. 2012; 17:1515–33. https://doi.org/10.1634/theoncologist.2012-0139 PMID: 23024155

24. Gilbert L, Basso O, Sampalis J, Karp I, Martins C, Feng J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. Lancet Oncol. 2012; 13:285–291. https://doi.org/10.1016/S1470-2045(11)70333-3 PMID: 22257524

25. Harrell FE. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.

26. Seed P. DIAGT: Stata module to report summary statistics for diagnostic tests compared to true disease status. Statistical Software Components S423401, Boston College Department of Economics. 2010.

27. Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. Research Data Supporting "The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study." University of Cambridge Repository; 2020. https://doi.org/10.17863/CAM.56363

28. Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case-control study. BMJ Open. 2016; 6:e011664. https://doi.org/10.1136/bmjopen-2016-011664 PMID: 27178981

29. Pauler DK, Menon U, McIntosh M, Symecko HL, Skates SJ, Jacobs IJ. Factors influencing serum CA125II levels in healthy postmenopausal women. Cancer Epidemiol Biomarkers Prev. 2001; 10:489–493. PMID: 11352859

30. Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. Accuracy of CA 125 in the diagnosis of ovarian tumors: A quantitative systematic review. Eur J Obstet Gynecol Reprod Biol. 2009; 142:99–105. https://doi.org/10.1016/j.ejogrb.2008.08.011 PMID: 18995946

31. Jacobs I, Bast RC. The CA 125 tumour-associated antigen: a review of the literature. Hum Reprod. 1989; 4:1–12.

32. Buys SS, Partridge E, Greene MH, Prorok PC, Reding D, Riley TL, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: Findings from the initial screen of a randomized trial. Am J Obstet Gynecol. 2005; 193: 1630–1639. https://doi.org/10.1016/j.ajog.2005.05.005 PMID: 16260202

33. Sjövall K, Nilsson B, Einhorn N. The significance of serum CA 125 elevation in malignant and nonmalignant diseases. Gynecol Oncol. 2002; 85:175–178. https://doi.org/10.1006/gyno.2002.6603 PMID: 11925140

34. Cancer Research UK. Ovarian cancer incidence by age. [cited 2020 Jun 5]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One

35. Banks J, Hollinghurst S, Bigwood L, Peters TJ, Walter FM, Hamilton W. Preferences for cancer investigation: A vignette-based study of primary-care attendees. Lancet Oncol. 2014; 15:232–240. https://doi.org/10.1016/S1470-2045(13)70588-6 PMID: 24433682

36. Reade CJ, Riva JJ, Busse JW, Goldsmith CH, Elit L. Risks and benefits of screening asymptomatic women for ovarian cancer: A systematic review and meta-analysis. Gynecol Oncol. 2013; 130:674–681. https://doi.org/10.1016/j.ygyno.2013.06.029 PMID: 23822892

37. Sundar S, Rick C, Dowling F, Au P, Snell K, Rai N, et al. Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. BMJ Open. 2016; 6:e010333. https://doi.org/10.1136/bmjopen-2015-010333 PMID: 27507231

38. Kurtz AB, Tsimikas J V., Tempany CMC, Hamper UM, Arger PH, Bree RL, et al. Diagnosis and Staging of Ovarian Cancer: Comparative Values of Doppler and Conventional US, CT, and MR Imaging Correlated with Surgery and Histopathologic Analysis—Report of the Radiology Diagnostic Oncology Group. Radiology. 1999; 212:19–27. https://doi.org/10.1148/radiology.212.1.r99jl3619 PMID: 10405715

39. Lee ES, Lee JM. Imaging diagnosis of pancreatic cancer: A state-of-the-art review. World J Gastroenterol. 2014; 20:7864–77. https://doi.org/10.3748/wjg.v20.i24.7864 PMID: 24976723

40. Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. Results of the Two Incidence Screenings in the National Lung Screening Trial. N Engl J Med. 2013; 369:920–931. https://doi.org/10.1056/NEJMoa1208962 PMID: 24004119

41. Menon U, Ryan A, Kalsi J, Gentry-Maharaj A, Dawnay A, Habib M, et al. Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. J Clin Oncol. 2015; 20:2062–71.

cancers

MDPI

Review

Identifying Ovarian Cancer in Symptomatic Women: A Systematic Review of Clinical Tools

Garth Funston [1,*], Victoria Hardy [1], Gary Abel [2], Emma J. Crosbie [3,4], Jon Emery [1,5], Willie Hamilton [2] and Fiona M. Walter [1,5]

1 The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; veh29@medschl.cam.ac.uk (V.H.); jon.emery@unimelb.edu.au (J.E.); fmw22@medschl.cam.ac.uk (F.M.W.)

2 University of Exeter Medical School, University of Exeter, Exeter EX1 1TX, UK; G.A.Abel@exeter.ac.uk (G.A.); W.Hamilton@exeter.ac.uk (W.H.)

3 Gynaecological Oncology Research Group, Division of Cancer Sciences, University of Manchester, Manchester M13 9WL, UK; emma.crosbie@manchester.ac.uk

4 Department of Obstetrics and Gynaecology, Manchester University NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester M13 9WL, UK

5 Centre for Cancer Research and Department of General Practice, University of Melbourne, Melbourne, VIC 3000, Australia

* Correspondence: gf272@medschl.cam.ac.uk

Received: 12 November 2020; Accepted: 4 December 2020; Published: 8 December 2020

Simple Summary: Most women with ovarian cancer are diagnosed after they develop symptoms—identifying symptomatic women earlier has the potential to improve outcomes. Tools, ranging from simple symptom checklists to diagnostic prediction models that incorporate tests and risk factors, have been developed to help identify women at increased risk of undiagnosed ovarian cancer. In this review, we systematically identified studies evaluating these tools and then compared the reported diagnostic performance of tools. All included studies had some quality concerns and most tools had only been evaluated in a single study. However, four tools were evaluated in multiple studies and showed moderate diagnostic performance, with relatively little difference in performance between tools. While encouraging, further large and well-conducted studies are needed to ensure these tools are acceptable to patients and clinicians, are cost-effective and facilitate the early diagnosis of ovarian cancer.

Abstract: In the absence of effective ovarian cancer screening programs, most women are diagnosed following the onset of symptoms. Symptom-based tools, including symptom checklists and risk prediction models, have been developed to aid detection. The aim of this systematic review was to identify and compare the diagnostic performance of these tools. We searched MEDLINE, EMBASE and Cochrane CENTRAL, without language restriction, for relevant studies published between 1 January 2000 and 3 March 2020. We identified 1625 unique records and included 16 studies, evaluating 21 distinct tools in a range of settings. Fourteen tools included only symptoms; seven also included risk factors or blood tests. Four tools were externally validated—the Goff Symptom Index (sensitivity: 56.9–83.3%; specificity: 48.3–98.9%), a modified Goff Symptom Index (sensitivity: 71.6%; specificity: 88.5%), the Society of Gynaecologic Oncologists consensus criteria (sensitivity: 65.3–71.5%; specificity: 82.9–93.9%) and the QCancer Ovarian model (10% risk threshold—sensitivity: 64.1%; specificity: 90.1%). Study heterogeneity precluded meta-analysis. Given the moderate accuracy of several tools on external validation, they could be of use in helping to select women for ovarian cancer investigations. However, further research is needed to assess the impact of these tools on the timely detection of ovarian cancer and on patient survival.

Keywords: ovarian cancer; symptoms; early detection; risk assessment; diagnostic prediction model; triage tool; ovarian cancer symptoms

1. Introduction

Ovarian cancer is the eighth most common cancer to affect women worldwide, accounting for over 384,000 deaths in 2018 [1]. Outcomes are strongly linked to stage at diagnosis, with five-year survivals of 90% and 4% for UK women diagnosed at stages I and IV, respectively [2]. Given this, large ovarian cancer screening trials have been conducted, but these have so far failed to demonstrate a significant reduction in long-term mortality [3,4]. In the absence of effective screening programs, the majority of ovarian cancers are diagnosed following symptomatic presentation [5,6], and a focus has been placed on the early detection of symptomatic disease [7].

While once regarded as a 'silent killer', many studies have demonstrated that a range of symptoms are more common in women with ovarian cancer than in control subjects and that symptoms occur at all stages of the disease [8]. Clinical guidelines in countries around the world recommend that patients presenting with symptoms of possible ovarian cancer undergo investigation, although debate remains over which symptoms are indicative of disease and should be included in guidelines [7]. To facilitate the early detection of symptomatic cancer, researchers have developed a number of symptom-based checklists for use either when patients first present in the clinical setting or in 'symptom-triggered screening' programs, in which symptoms are proactively solicited [9–11]. More sophisticated tools, which can take the form of diagnostic prediction models [12], have also been developed to incorporate test results and ovarian cancer risk factors alongside symptoms, in a bid to improve tool performance. Several of these tools have been incorporated into clinical computer systems, which, then, automatically alert the clinician to consider ovarian cancer investigations when relevant symptoms are present or when the risk of undiagnosed cancer reaches a certain level. However, the relative limitations and merits of the various available tools remain unclear. In this systematic review, we aimed to identify and compare the diagnostic performances of symptom-predicated tools for the detection of ovarian cancer.

2. Methods

2.1. Eligibility Criteria and Searches

This review was conducted and is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Table S1); a study protocol was registered with PROSPERO [CRD42020149879]. We searched MEDLINE, EMBASE and Cochrane CENTRAL for keywords relating to ovarian cancer, symptoms and prediction/diagnostic tools to identify papers published between 1 January 2000 and 3 March 2020 (Text S1). The start date was chosen to predate the publication of key ovarian cancer symptom papers [13,14]. No language restrictions or restrictions on methodological design were applied. No restrictions were placed on study setting, so studies conducted in the general population or in primary, secondary, or tertiary care were all eligible for inclusion. Reference lists of included papers were screened to identify any additional relevant papers.

Studies were included if they (a) described the development and or evaluation of a multivariable tool designed to identify patients with undiagnosed ovarian cancer and (b) provided the sensitivity and specificity of the tool or gave sufficient data to allow these metrics to be calculated. For the purposes of this review, we defined a multivariable tool as a combination of three or more variables used to detect or predict the risk of undiagnosed ovarian cancer. This broad definition encompasses traditional multivariable diagnostic prediction models and clinical prediction rules [12,15]. We considered variable 'checklists', in which any one variable in the list needed to be present for a positive result, to be a form of multivariable tool. As the focus of this review was on symptom-based tools, the tool under investigation had to include at least one symptom for a study to be eligible. No other restrictions were placed on the type of variable that could be included in a tool. Studies on tools intended to estimate future risk of developing ovarian cancer rather than the current risk of having an undiagnosed ovarian cancer were excluded, as were studies on tools that solely provide an indication of the risk of relapse or recurrence. We excluded studies in which all participants had a pelvic mass—as this represents a highly

selected high-risk population—and studies undertaken solely in paediatric (<18 years) populations. Non-primary research studies were also excluded.

2.2. Study Selection

The online Rayyan software was used to facilitate abstract screening and study selection [16]. Following removal of duplicates, two reviewers (G.F. and V.H.) independently screened titles and abstracts against eligibility criteria. Potentially eligible papers identified at the screening stage were obtained and the full texts were independently examined against eligibility criteria by two reviewers (G.F. and V.H.). Any disagreements were resolved by consensus.

2.3. Data Extraction and Synthesis

Data extraction was performed by one reviewer (G.F.) and checked against full-text papers by a second reviewer (V.H.) to ensure accuracy. Using a predeveloped template, information was extracted on study characteristics (year of publication and location), study design (methodology, population, data source and outcome definition), tools (variables and tool development methods), and tool performance metrics (sensitivity, specificity and other diagnostic metrics). Where a study evaluated multiple tools, data relating to each tool were extracted separately.

Sensitivity and specificity were used to compare tool accuracy. For diagnostic prediction models, area under the receiver operator characteristic curve (AUC) was used to compare discrimination (the ability of a tool to identify those with a condition from those without a condition) and calibration (agreement between estimated and observed outcomes). Due to the marked heterogeneity of included studies in terms of the study designs, populations, variable definitions, outcome definitions and use of different tool thresholds, and the failure of multiple studies to report numbers of patients with true positive/true negative/false positive/false negative results, we were unable to perform any meta-analyses. Instead, performance characteristics were summarised in tabular form and using a narrative synthesis approach. When synthesising data, we paid particular attention to several study and tool characteristics. First, the source of participant recruitment. For example, whether controls were recruited from the general population or after entry into healthcare, as symptoms may be more common in clinical controls than population controls, which could influence measures of tool sensitivity and specificity [17]. Second, whether the measures of tool accuracy were obtained directly from the patient sample in which the tool was developed (apparent performance), by applying internal validation methods, such as splitting the sample into development and validation sets or using cross-validation techniques (internal validation), or from a separate analysis in a distinct population (external validation) [12]. Tools usually exhibit poorer diagnostic performance in external validation studies than when evaluated in the original development sample, and external validation of tools is recommended before they are used in clinical practice [12]. Third, we considered whether tools consisted solely of symptoms or symptoms in addition to other variables, as this is likely to impact the clinical utility of the tool.

2.4. Risk of Bias Assessment

The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool was used to assess the risk of bias and applicability of the included studies [18]. QUADAS-2 includes signalling questions (intended to identify areas of potential bias or concern over study applicability) covering four domains: (1) patient selection, (2) index test(s), (3) reference standard and (4) flow and timing. Each domain was rated as having "high", "low" or "unclear" (where insufficient information is provided) risk of bias. Domains 1–3 were also rated for applicability as "high", "low" or "unclear" concern. Two reviewers (G.F. and V.H.) independently assessed each study using QUADAS-2. Ratings were compared and disagreements were resolved by consensus.

3. Results

3.1. Study Selection

In total, 2331 records were identified from database searches, of which 708 were duplicates. Two additional records were identified from examination of reference lists. A total of 1625 titles and abstracts were screened, and 35 full-text papers were examined. Sixteen studies met the eligibility criteria and were included (Figure 1).

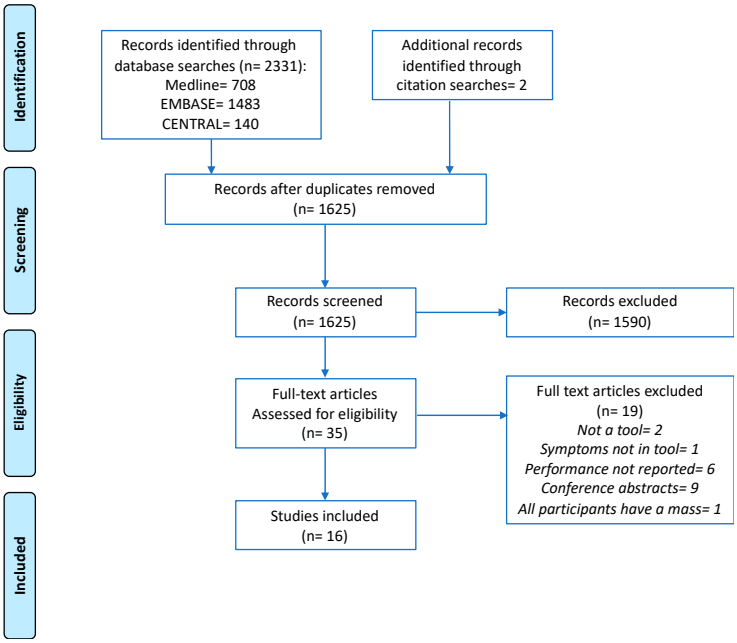


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram illustrating the study selection process.

3.2. Study Characteristics

The characteristics of the included studies are summarised in Table 1 and additional exclusion criteria are detailed in Supplementary Material Table S2. Three studies were population-based [19–21], five studies were based in a primary care setting [14,22–25], four studies were entirely hospital-based [26–29] and four studies were hospital-based but also recruited controls from screening studies [30–33]. All population- and hospital-based studies were of case-control design. Two of the studies that recruited from the hospital setting included a proportion of controls with benign ovarian pathology [26,28]. Three of the five primary care studies were of cohort design [22–24], and the remaining two were of case-control design [14,25]. The studies used a variety of data sources for variables, including pre-existing routinely collected primary care data (n = 6), information from surveys or patient interviews (n = 11) and blood samples (n = 4). Study sizes varied markedly, with 75–1,908,467 participants and 24–1885 women with ovarian cancer per study. While all studies used ovarian cancer as an outcome, how this was defined differed, with some only including invasive epithelial cancer or specifically stating that they excluded borderline tumours [19–21,26–29], and others apparently including both invasive and borderline epithelial tumours or all ovarian cancers [14,22–25,30–33]. One study included ovarian cancer alongside other common cancers in a composite outcome, but tool performance characteristics for each cancer were given separately [23]. Seven studies developed entirely new tools [14,19,22,23,25,30,33], six modified existing tools [26–29,31,32] and eight externally validated existing tools [20,21,24,26–29,33].

Table 1. Study characteristics.

Author, Date, Country	Design		Objective			Primary Outcome	Candidate Variable Data Sources	Participants	Study Sample
	Case Control	Cohort	Develop New Tool	Modify Existing Tool	Externally Validate Existing Tool				
Population based									
Lurie, 2009, USA	•		•			Primary invasive ovarian carcinoma	In-person patient interviews using a structured survey	Cases: Women aged 19–88 years histologically-confirmed primary invasive ovarian carcinoma (1993–2007) Controls: Aged ≥ 18 years, Hawaii resident ≥ 1 year, randomly selected from statutory state survey Frequency-matched to cases (1:1) by age, ethnicity, interview time	Cases: 432 Controls: 491
Rossing, 2010, USA	•				•	Primary invasive epithelial OC [a]	In-person interviews	Cases: Residents in western Washington State, aged 35–74 years, diagnosed with a primary invasive epithelial ovarian tumour (January 2002–December 2005) Controls: Selected by random digit dialling with stratified sampling in 5-year age categories, 1-year calendar intervals and two (urban vs. suburban or rural) county strata	Cases: 594 Controls: 1313
Jordan, 2010, Australia	•				•	Invasive epithelial OC	Patient survey	Cases: Aged 20–79 years with suspected OC, subsequently diagnosed with invasive epithelial OC (January 2002–June 2005) Controls: Frequency-matched based on age (5-year groups) and state of residence identified from electoral roll [34]	Cases: 1215 Controls: 1456
Primary care population									
Hamilton, 2009, England	•		•			Primary OC, including borderline	Researcher-coded GP records	Cases: Aged ≥ 40 years with primary OC diagnosed between 2000 and 2007 Controls: Matched on age, sex and GP practice	Cases: 212 Controls: 1060
Hippisley-Cox, 2012, England and Wales		•	•			OC (NOS)	QResearch database [35]	Aged 30–84 years, registered with GP practices between 1 January 2000 and 30 September 2010	Development (2/3)—1,158,723 women with 976 OCs Validation (1/3)—608,862 women with 538 OCs

Table 1. *Cont.*

Author, Date, Country	Design		Objective			Primary Outcome	Candidate Variable Data Sources	Participants	Study Sample
	Case Control	Cohort	Develop New Tool	Modify Existing Tool	Externally Validate Existing Tool				
Hippisley-Cox, 2013, England and Wales		•	•			OC (NOS) and 10 other cancers	QResearch database [35]	Aged 25–89 years, registered with GP practices between 1 January 2000 and 1 April 2012	Development (2/3)—1,240,864 women with 1279 OCs Validation (1/3)—667,603 women with 606 OCs
Grewal, 2013, England	•		•			Primary OC, including borderline	Researcher-coded GP records	Cases: Aged ≥ 40 years with primary OC diagnosed between 2000 and 2007 Controls: Matched on age, sex and GP practice	Cases: 212 Controls: 1060
Collins, 2013, UK		•			•	OC (NOS)	THIN database [36]	Women aged 30–84 years registered with GP practices between 1 January 2000 and 30 June 2008	1,054,818 women with 735 cancers
Hospital + screening populations									
Goff, 2007, USA	•		•			OC, including borderline	Patient survey	Cases: Women with a pelvic mass recruited in secondary care prior to OC diagnosis Controls: (a) Healthy 'high-risk' [b] women enrolled in a screening study [37], (b) women who presented for pelvic/abdominal US	Development Cases: 74 Controls: 243 Validation Cases: 75 Controls: 245
Andersen, 2008, USA	•			•		OC (NOS)	Patient survey, blood sample	Cases: Women with a pelvic mass, recruited prior to OC diagnosis Controls: Healthy 'high risk' [b] women enrolled in a screening study [37]	Cases: 75 Controls: 254
Andersen, 2010, USA	•			•		OC (NOS)	Patient survey, blood sample	Cases: Women with a pelvic mass recruited in secondary care prior to OC diagnosis Controls: Healthy 'high risk' [b] women enrolled in a screening study [37], frequency matched to cases on age (</>50 years)	Cases: 74 Controls: 137
Lim, 2012, UK	•		•		•	OC, including borderline	(a) Survey, (b) telephone interview, (c) GP notes	Cases: Women aged 50–79 years with primary OC recruited prior to diagnosis (February 2006–February 2008) Controls: Screening trial participants [38], frequency matched on year of birth and agreement to a telephone interview	Cases: 194 [c] Controls: 268 [c]

Table 1. *Cont.*

Author, Date, Country	Design		Objective			Primary Outcome	Candidate Variable Data Sources	Participants	Study Sample
	Case Control	Cohort	Develop New Tool	Modify Existing Tool	Externally Validate Existing Tool				
Hospital based population									
Kim, 2009, Korea	•			•	•	Epithelial OC (NOS)	Patient survey, blood sample	Cases: OC diagnosis Controls: Women with benign ovarian cysts recruited prior to surgery and those undergoing routine pap smear	Cases: 116 Controls: 209 (Benign: 74, Pap smear: 135)
Macuks, 2011, Latvia	•			•	•	Epithelial OC (NOS)	Patient survey, blood sample	Cases: Women with epithelial OC recruited prior to surgery/diagnosis Controls: Age-matched 'healthy women' attending a gynaecology outpatient clinic [d]	Cases: 24 Controls: 31 [d]
Shetty, 2015, India	•			•	•	OC, excluding borderline	Patient survey	Cases: Women admitted to hospital for investigation and subsequently diagnosed with OC Controls: (a) Women with benign ovarian pathology; (b) those undergoing a 'gynaecological check-up'	Cases: 74 Controls: 218 (benign: 144, gynaecological check-up: 74)
Jain, 2018, India	•			•	•	OC, excluding borderline	Patient survey, blood sample	Cases: Women undergoing surgery for a pelvic mass, subsequently diagnosed with ovarian cancer Controls: First-degree healthy relatives of cases	Cases: 45 Controls: 90

[a] Data collected on borderline tumours but not included in their tool performance evaluation. [b] Women with high-risk family histories consistent with a possible BRCA1/2 mutation in their families, participating in the Ovarian Cancer Early Detection Study (OCEDS) [37]. [c] Numbers varied by study component: questionnaire (191 cases, 268 controls), telephone interview (111 cases, 125 controls) and GP notes (171 cases, 227 controls). [d] Controls with benign gynaecological disease were also included in study but are excluded from the review, as performance was examined separately to healthy controls and no overall specificity measure was given. Study design and Objectives denoted by "•". Abbreviations: OC = Ovarian cancer; NOS = Not otherwise specified; GP = General practice; US = Ultrasound.

3.3. Risk of Bias

The main potential sources of bias were identified in the "patient selection" and the "index test" domains (Figure 2). As the case-control design can lead to overestimation of test performance [18], 13 studies were flagged as being at high risk of bias for patient selection. Key potential sources of bias identified for studies in the "index test" domain included failing to pre-define the tool threshold and retrospectively administering the tool after the outcome had been determined, e.g., questioning participants after the ovarian cancer diagnosis had been made. The risk of bias was generally judged as low for the "reference standard" and "flow and timing" domains. However, all primary care studies were flagged as being at high risk of bias in the "reference standard" domain as they relied on general practitioner (GP) records to identify ovarian cancer diagnoses, supplemented in two studies by death registration data [22,23] rather than hospital or cancer registry histological diagnoses. Concern over the applicability of studies was judged as low, save for the "reference standard" domain of one study which used a composite cancer outcome [23].

Study	Risk of bias domain			
	Patient selection	Index test	Reference standard	Flow and timing
Lurie, 2009				
Rossing, 2010				
Jordan, 2010				
Hamilton, 2009				
Hippisley-Cox, 2012				
Hippisley-Cox, 2013				
Grewal, 2013				
Collins, 2013				
Goff, 2007				
Andersen, 2008				
Andersen, 2010				
Lim, 2012				
Kim, 2009				
Macuks, 2011				
Shetty, 2015				
Jain, 2018				

Figure 2. QUADAS-2 Risk of Bias Assessment. Green = "low", orange = "high", blue = "unclear" risk of bias.

4. Tool Variables

The studies evaluated a total of 21 distinct tools, of which five were diagnostic prediction models developed using appropriate statistical methods from which variable weights were derived [12]. We grouped variables included in the tools into four categories: (1) patient demographics, (2) personal and family medical history, (3) symptoms and (4) test results (Table 2). By definition, all tools included symptoms, with 14 including only symptoms. Four tools incorporated demographics, two incorporated personal and family medical history and six incorporated test results. Five symptoms (abdominal pain, pelvic pain, distension, bloating and appetite loss) were included in more than half (≥11) of the tools and a further six symptoms (feeling full quickly, difficulty eating, postmenopausal bleeding, urinary frequency, palpable abdominal mass/lump and rectal bleeding) were included in at least a quarter (≥6) of the tools. Six tools were based on an existing tool—the Goff Symptom Index (SI)—which was modified to include additional symptom or test result variables. Specifications of each tool, including how variables were defined, are included in the Supplementary Material Table S3.

Table 2. Variables included in the final tools.

Tool (Study, Year)	Demographics		Personal/Family History		Symptoms												Test Results		
	Age	Other	PMH	FH	Abdo. Pain	Pelvic Pain	Increase Abdo. Size/Distens.	Bloat.	Appetite Loss	Feeling Full	Difficulty Eating	Weight Loss	Postmen. Bleeding	Rectal Bleeding	Palpable Abdo. Mass/lump	Urinary Freq.	Other	Hb	CA125	HE4
Symptom checklists																				
Goff SI (Goff, 2007)					•	•	•	•		•	•									
Modified Goff SI 1 (Kim, 2009)					•	•	•	•		•	•					•	Urinary urgency			
Lurie 7-SI (Lurie, 2009					•	•	•	•	•						•	•	Bowel symptoms, difficulty emptying bladder, dysuria, fatigue, abnormal vaginal bleed.			
Lurie 5-SI (Lurie, 2009)					•	•	•								•	•	Difficulty emptying bladder, dysuria, abnormal vaginal bleed.			
Lurie 4-SI (Lurie, 2009)					•	•	•								•		Abnormal vaginal bleed.			
Lurie 3-SI (Lurie, 2009)							•								•		Abnormal vaginal bleed.			
Hamilton SI (Hamilton, 2009)					•		•	•	•				•	•		•				
SGO consensus criteria * (Rossing, 2010)					•	•		•		•						•	Urinary urgency			
Lim SI 1 (Lim, 2012)					•	•	•	•	•	•		•			•					
Lim SI 2 (Lim, 2012)					•	•	•		•						•		Vaginal discharge			
Hippisley-Cox SI (Hippisley-Cox, 2012)					•		•		•			•	•	•						

Table 2. *Cont.*

Tool (Study, Year)	Demographics		Personal/Family History		Symptoms												Test Results		
Modified Goff SI 2 (Shetty, 2015)					•	•	•	•	•	•	•	•			•	Urinary urgency			
	Age	**Other**	**PMH**	**FH**	**Abdo. Pain**	**Pelvic Pain**	**Increase Abdo. Size/Distens.**	**Bloat.**	**Appetite Loss**	**Feeling Full**	**Difficulty Eating**	**Weight Loss**	**Postmen. Bleeding**	**Rectal Bleeding**	**Palpable Abdo. Mass/lump**	**Urinary Freq.**	**Other**	**Hb**	**CA125**	**HE4**
Augmented symptom checklists																			
Goff SI + CA125 (Andersen, 2008)					•	•	•	•		•	•								•	
Goff SI + HE4 (Andersen, 2010)					•	•	•	•		•	•									•
Goff SI + HE4 + CA125 (Andersen, 2010)					•	•	•	•		•	•								•	•
Goff SI + CA125 + menopause (Macuks, 2011)		Menopause			•	•	•	•		•	•								•	
Prediction models																				
QCancer Ovarian (Hippisley-Cox, 2012)	•			OC	•		•		•			•	•	•				•		
QCancer Female (Hippisley-Cox, 2013)	•	Townsend score, smoking, alcohol, BMI	T2DM, COPD, endomet. hyperplasia or polyp, chronic pancreatitis	OC, GI cancer, breast cancer	•		•		•			•	•	•			Difficulty swallowing, heartburn/indigestion, blood in urine, blood in vomit, blood when cough, irregular menstrual bleeding, vaginal bleeding after sex, breast lump, breast skin tethering or nipple discharge, breast pain, lump in neck, night sweats, venous thromboembolism, CIBH, constipation, cough, unexplained bruising	•		

Table 2. *Cont.*

Tool (Study, Year)	Demographics		Personal/Family History		Symptoms												Test Results		
	Age	Other	PMH	FH	Abdo. Pain	Pelvic Pain	Increase Abdo. Size/Distens.	Bloat.	Appetite Loss	Feeling Full	Difficulty Eating	Weight Loss	Postmen. Bleeding	Rectal Bleeding	Palpable Abdo. Mass/lump	Urinary Freq.	Other	Hb	CA125	HE4
OC Score A (Grewal, 2013)					•		•	•	•				•	•		•				
OC Score B (Grewal, 2013)					•		•	•	•				•	•		•				
OC Score C (Grewal, 2013)	•				•		•	•	•				•	•		•				

* Consensus statement released by the Society of Gynaecologic Oncologists (SGO), the Gynaecologic Cancer Foundation and the American Cancer Society. The presence of a variable within a model is denoted by "•". The terms used to describe a given symptom varied subtly between studies—full details of each tool, including symptom terminology and duration and frequency criteria, are included in Supplementary Material Table S3. Abbreviations: PMH = past medical history; FH = family history; Abdo. = abdominal; Distens. = distension; Bloat. = bloating; Postmen. = postmenopausal; bleed. = bleeding; Freq. = frequency; Hb = haemoglobin; CA125 = cancer antigen 125; HE4 = human epididymis protein 4; SI = symptom index; OC = ovarian cancer; BMI = body mass index; endomet. = endometrial; T2DM = type 2 diabetes mellites; COPD = chronic obstructive pulmonary disease; GI = gastrointestinal; CIBH = change in bowel habit.

4.1. Evaluation of Tool Performance

The diagnostic performance of the included tools is summarised in Table 3. Measures of diagnostic performance for the majority of the tools were obtained directly from the patient sample with which the tool was developed (apparent performance) or by applying internal validation methods, such as splitting the sample into development and validation sets (internal validation), with only four tools—the Society of Gynaecologic Oncology (SGO) consensus criteria, Goff SI, QCancer Ovarian, Modified Goff SI 1—undergoing independent validation with an external dataset. Although the Goff SI in combination with CA125 was evaluated in several studies, the CA125 thresholds used varied markedly, so no studies were considered to have externally validated the same combination. There was overlap in evaluation of tools between healthcare settings, but no tool evaluated in primary care was evaluated in another setting or vice versa.

The most widely studied tool was the Goff SI, which was evaluated in nine studies [20,21,26,27,29–33], but two of these used data from subsets of women in the original tool development study [31,32]. Apparent deviations from the original Goff SI in how variables were defined were noted in several studies (Table S4). The Goff SI was the only tool to be externally validated in groups of women recruited from more than one setting.

4.2. Tool Diagnostic Accuracy

4.2.1. Hospital Setting

All but two tools evaluated in hospital populations incorporated the Goff SI. Two of these underwent external evaluation—the original Goff SI and a modified version incorporating additional symptoms (Modified Goff SI 1). The Goff SI, which was externally validated in six studies, demonstrated sensitivities which ranged from 56.9% to 83.3% (an outlier result) and specificities from 48.3% (an outlier result) to 98.9%. A modified version of the Goff SI (Modified Goff SI 1) demonstrated a sensitivity of 71.6% and a specificity of 88.5% in a single external validation study.

Augmenting symptom checklists with baseline risk factors and test results generally led to a reduction in sensitivity and an increase in specificity, or vice versa, depending on the threshold used. For example, the addition of the serum ovarian cancer biomarker CA125 to the Goff SI by Anderson et al. (2008) led to a reduction in tool sensitivity—if both variables were required to be abnormal for a positive tool result—or in tool specificity—if only one was required to be abnormal for a positive tool result [31].

4.2.2. Population Setting

In women recruited from the population setting, two symptom checklists were externally validated side by side—the Goff SI and the SGO consensus criteria. While the sensitivities and specificities of the tools differed between the studies, within each study, they were similar, with an in-study maximum difference in sensitivity of 3.4% and specificity of 2.4% between the tools.

4.2.3. Primary Care

A single tool (QCancer Ovarian), which took the form of a prediction model and combined symptom variables with demographics, family history and routine blood test results, underwent external validation in a primary care setting. When the threshold for abnormality was set to include the 5% of women at the highest predicted risk, QCancer Ovarian had a sensitivity of 43.8% and a specificity of 95%, while when the threshold was set to include women at the 10% highest risk, the sensitivity increased to 64.1% but the specificity fell to 90.1%. Several scores, developed by Grewal et al., demonstrated higher sensitivities and specificities than QCancer Ovarian at the 5% risk threshold (OC Score B $\geq$ 4) and 10% risk threshold (OC Score C $\geq$ 4), but diagnostic accuracy measures were derived from the same dataset used in score development.

Table 3. Tool diagnostic accuracy.

Tool	Study	Recruitment			Source of Accuracy Estimate				Sensitivity (95% CI)	Specificity (95% CI)	PPV	AUC (95% CI)
		Population Level	1° Care	Hospital + Screening	Hospital	Apparent Performance	Internal Validation	External Validation				
Symptom checklists												
Goff SI	Goff, 2007			•			•		≥50 yrs: 66.7 <50 yrs: 86.7	≥50 yrs: 90 <50 yrs: 86.7	-	-
	Andersen, 2008 [a]			•			•		64 (52.1–74.8)	88.2 (83.6–91.9)	-	-
	Kim, 2009				•			•	56.9	87.6	-	-
	Rossing, 2010	•						•	67.5 (65.4–69.6)	94.9 (93.9–95.8)	0.77–1.12 [b]	-
	Jordan, 2010	•						•	68.1 (65.5–70.7)	85.3	0.09 [c] (≥55 yrs: 0.21–0.31 <55 yrs: 0.04) [d]	-
	Andersen, 2010 [a]			•			•		63.5 (51.5–74.4)	88.3 (81.7–93.2)	-	-
	Macuks, 2011				•			•	83.3	48.3	-	-
	Jain, 2018				•			•	77.8	87.8	-	-
	Lim, 2012			•				•	61.4–75.7 [e]	89.6–98.9 [e]	-	-
Modified Goff SI 1	Kim, 2009				•	•			65.5	84.7	-	-
	Shetty, 2015				•			•	71.6	88.5	-	-
7-symptom Index	Lurie, 2009	•				•			85	40	-	-
5-symptom Index	Lurie, 2009	•				•			80	63	-	-
4-symptom Index	Lurie, 2009	•				•			74	77	-	-
3-symptom Index	Lurie, 2009	•				•			54	93	-	-
Hamilton SI	Hamilton, 2009		•			•			85	85	-	-
SGO consensus criteria	Rossing, 2010	•						•	65.3 (63.1–67.4)	93.9 (92.8–95)	0.63–0.92 [b]	-
	Jordan, 2010	•						•	71.5 (69–74.1)	82.9 (81–84.8)	0.08 [c] (≥55 yrs: 0.18–0.27 <55 yrs: 0.05) [d]	-
Lim SI 1	Lim, 2012			•			•		69.6–91 [e]	76–91 [e]	-	-

Table 3. *Cont.*

Tool	Study	Recruitment			Source of Accuracy Estimate			Sensitivity (95% CI)	Specificity (95% CI)	PPV	AUC (95% CI)
		Population Level	1° Care	Hospital + Screening	Hospital	Apparent Performance	Internal Validation	External Validation				
Lim SI 2	Lim, 2012			•			•		67.3–91 [e]	82.4–94 [e]	-	-
Hippisley-Cox SI	Hippisley-Cox, 2012		•				•		71.9	82.9	0.5	-
Modified Goff SI 2	Shetty, 2015				•	•			77	88.5	-	-
Augmented symptom checklists												
Goff SI or CA125 [f]	Andersen, 2008			•		•			89.3 (80.1–95.3)	83.5 (78.3–87.8)	-	-
Goff SI or CA125 (>35 U/mL)	Jain, 2018				•	•			97.8	68.9	-	-
Goff SI & CA125 (>21 U/mL)	Macuks, 2011				•	•			79.1	100	-	-
Goff SI & CA125 (>35 U/mL)	Macuks, 2011				•	•			70.8	100	-	-
Goff SI & CA125 (>65 U/mL)	Macuks, 2011				•	•			70.8	100	-	-
Goff SI or CA125 [f]	Andersen, 2010			•		•			91.9 (83.2–97)	83.2 (75.9–89)	-	-
Goff SI or HE4 [f]	Andersen, 2010			•		•			91.9 (83.2–97)	84.7 (77.5–90.3)	-	-
Any 1 of 3 (Goff SI or CA125 or HE4) [f]	Andersen, 2010			•		•			94.6 (86.7–98.5)	79.6 (71.8–86)	-	-
Any 2 of 3 (Goff SI or CA125 or HE4) [f]	Andersen, 2010			•		•			83.8 (73.4–91.3)	98.5 (94.8–99.8)	-	-
Goff SI & 1 or more of CA125 or HE4 [f]	Andersen, 2010			•		•			58.1 (46.1–69.5)	98.5 (94.8–99.8)	-	-
Goff SI & CA125 (>25 U/mL) & menopause	Macuks, 2011				•	•			50	100	-	-
Goff SI & CA125 (>35 U/mL) & menopause	Macuks, 2011				•	•			45.8	100	-	-
Goff SI & CA125 (>65 U/mL) & menopause	Macuks, 2011				•	•			45.8	100	-	-
Prediction models												

Table 3. *Cont.*

Tool	Study	Recruitment			Source of Accuracy Estimate				Sensitivity (95% CI)	Specificity (95% CI)	PPV	AUC (95% CI)
		Population Level	1° Care	Hospital + Screening	Hospital	Apparent Performance	Internal Validation	External Validation				
QCancer Ovarian (Top 10% risk)	Hippisley-Cox, 2012		•				•		63.2	90.8	0.8	084 (0.83–0.86)
	Collins, 2013		•					•	64.1	90.1	0.5	0.86 (0.84–0.87)
QCancer Ovarian (Top 5% risk)	Hippisley-Cox, 2012		•				•		42.2	95.6	1.1	-
	Collins, 2013		•					•	43.8	95	0.6	-
QCancer Ovarian (Top 1% risk)	Hippisley-Cox, 2012		•				•		13.9	99.3	2.1	-
QCancer Ovarian (Top 0.5% risk)	Hippisley-Cox, 2012		•				•		11	99.6	3.2	-
QCancer Ovarian (Top 0.1% risk)	Hippisley-Cox, 2012		•				•		3.9	99.9	5.5	-
QCancer Female (Top 10% risk)	Hippisley-Cox, 2013		•				•		61.6	90	0.6	0.84 (0.82–0.86)
OC Score A (Score ≥ 3)	Grewal, 2013		•			•			58.5	97.3	-	0.89
OC Score A (Score ≥ 4)	Grewal, 2013		•			•			57.6	97.3	-	
OC Score B (Score ≥ 3)	Grewal, 2013		•			•			75	90.1	-	0.89
OC Score B (Score ≥ 4)	Grewal, 2013		•			•			58.9	97.3	-	
OC Score C (Score ≥ 3)	Grewal, 2013		•			•			85.4	85.1	-	0.88
OC Score C (Score ≥ 4)	Grewal, 2013		•			•			72.6	91.3	-	

[a] Study used a subset of patients from Goff, 2007. [b] Calculated using external data from screening studies. [39,40]. [c] Calculated using external Australian population-level data. [d] Calculated using external data from US and UK screening studies and Australian population-level data. [41,42]. [e] Sensitivity and specificity varied by data collection method (questionnaire, telephone interview, GP notes). [f] Biomarker level (CA125, HE4) dichotomised at 95th percentile in control group—levels above that deemed abnormal. The Recruitment setting and the source of accuracy estimate are denoted by "•". Abbreviations: OC = ovarian cancer; CI = confidence interval; AUC = area under the receiver operator characteristic curve; PPV = positive predictive values; yrs = years.

Discrimination was reported for five tools (Table 3), all of which had similar AUCs within the 'good' range (0.84–0.89), with QCancer Ovarian exhibiting an AUC of 0.86 on external validation. Tool calibration was assessed for QCancer tools by graphically comparing the predicted cancer risk at two years with the observed risk by predicted risk deciles [22–24]. Authors reported good calibration on internal validation. On external validation, QCancer Ovarian had reasonable calibration but overpredicted risk, particularly in older women [24].

4.2.4. Positive Predictive Values

The three cohort studies conducted in primary care reported positive predictive values (PPV) for QCancer tools at a range of thresholds (Table 3). The PPVs at any given risk threshold were similar—for example, values ranged from 0.5 to 0.8% when the threshold was set to identify the 10% of women at highest risk. Two case control studies (Rossing et al. and Jordan et al.) used external disease prevalence figures from screening studies and available population-level statistics to estimate the PPVs of the Goff SI and SGO consensus criteria—if they were to be used in general populations. The tools had similar estimated PPVs within each study, but PPVs were higher in Rossing et al. (0.63–1.12%) than in Jordan et al. (<55 years: 0.04–0.05%, ≥55 years: 0.18–0.31%).

5. Discussion

To our knowledge, this is the first systematic review to compare the diagnostic performance of existing symptom-based tools for ovarian cancer detection. We identified 21 symptom-based tools designed to help identify women with undiagnosed ovarian cancer. These tools comprised simple symptom checklists, checklists which included both symptoms and tests and more complex diagnostic prediction models which incorporated symptoms, test results and baseline risk factors. While the diagnostic performances of most tools were evaluated solely within the study development datasets, four tools were independently externally validated, with one being validated in multiple population settings. Externally validated tools demonstrated similar moderate diagnostic performances. Our findings should inform future studies evaluating the clinical impact of validated symptom-based tools when implemented in clinical practice.

5.1. Study Strengths and Limitations

The main strengths of this study were its systematic approach, broad search strategy and liberal eligibility criteria, which enabled us to identify and compare the performances of a wide variety of tools. However, the identified studies were extremely heterogeneous in their designs, populations, variable definitions, outcome definitions and thresholds, which ultimately precluded any meaningful meta-analyses. For example, although the Goff SI was evaluated in nine studies, there was overlap between the participants in three studies, control groups ranged from apparently healthy general population participants to hospital gynaecology patients (with or without benign pathology), ovarian cancer definitions differed and deviations in the parameters of the SI itself, in terms of symptom duration and frequency criteria, were noted in several studies. While meta-analysis was not deemed appropriate, our results demonstrate how the Goff SI performs under different conditions. An additional limitation was that all included studies were at high risk of bias in at least one QUADAS-2 domain, which limits the conclusions that can be drawn.

5.2. Comparison of Tools

Although all tools were symptom-based and designed to help identify women with ovarian cancer, they varied markedly in the symptoms they included. This mirrors discrepancies in the literature and within national guidelines as to which symptoms are associated with the disease and probably reflects differences in study methodologies and study populations [7]. Despite this, the symptoms with the highest positive likelihood ratios for ovarian cancer in a recent systematic review (distension, bloating, abdominal or pelvic pain) were incorporated into the majority of tools [8]. The more cancer-associated symptoms that are included in a checklist, the higher the sensitivity of the tool is likely to be, but at the

cost of reducing specificity, as demonstrated by several of the included studies [19,26,33]. This was cited by Goff et al. as a rationale for not including urinary symptoms in the Goff SI [30]. Ultimately, variation in which additional symptoms a tool includes may have limited impact on tool performance; on external validation, two studies reported similar diagnostic accuracy metrics for the Goff SI and the SGO criteria (which differed on several symptoms), and on internal validation, Lim et al. concluded that changing several of the symptoms made relatively little difference to tool diagnostic accuracy [33].

In multiple studies, symptom checklists were augmented by ovarian cancer biomarkers with the aim of improving tool diagnostic accuracy. This approach naturally led to a reduction in tool specificity (where either symptoms or an abnormal test resulted in a positive tool) or sensitivity (where symptoms and an abnormal test were needed for a positive tool). If ovarian cancer biomarkers are to be included alongside symptoms within tools, this loss of performance could be avoided by incorporating them within prediction models, as per the inclusion of anaemia in QCancer Ovarian. As the prediction model threshold can be set at a desired risk level, biomarkers, such as CA125 and HE4, could be incorporated without harming tool performance. However, this would require women to have specialist ovarian cancer markers performed in order for the tool to be used, which significantly limits clinical utility. A more practical approach would be to incorporate tools within a two-step pathway in which symptom-based tools (which do not include specialist test variables) are used to help select higher-risk women for specialist ovarian cancer tests.

Variation in the reported sensitivity and specificity of the most widely evaluated tool, the Goff SI, was noted between studies. This variation is likely to be due, in part, to the marked differences in study design, populations and outcome definitions which precluded meta-analysis across these studies. Despite these differences, in 5 of the 6 external validation studies (including two large population-based studies), the Goff SI had a sensitivity in excess of 60%, and in all but the smallest study, which included only 24 ovarian cancers and 31 controls, its specificity exceeded 85%. The sensitivities and specificities of the two other externally validated symptom checklists—the SGO consensus criteria and the modified Goff SI 1—were similar, as were those of the only externally validated diagnostic prediction model—QCancer Ovarian (applying a 10% risk threshold). Given the similarity in performance of the various existing validated tools, future research efforts may be better directed at evaluating the impact of using available tools in practice rather than developing further tools consisting of different symptom combinations.

5.3. Clinical Relevance

Two distinct uses for tools were identified by the authors of the included studies: (1) assessment of women presenting symptomatically in the standard clinical setting to identify those at higher risk of undiagnosed cancer and to inform decision making and further investigation, and (2) proactive 'symptom-triggered screening' programs in which women are actively screened using the tool, with further testing for ovarian cancer occurring if the tool is positive. Several of the tools identified in this review are already available for use within the standard clinical setting in the form of electronic clinical decision support tools (eCDSTs). QCancer tools are integrated within some UK general practice IT systems and alert the clinician if the risk of ovarian cancer in an individual reaches a certain level, prompting them to consider ovarian cancer as a possible diagnosis. eCDSTs have been shown to improve practitioner performance and patient care, but there are multiple barriers to their implementation and they do not always lead to improved outcomes [43,44]. Therefore, even if eCDSTs are deemed to have acceptable diagnostic accuracy, their cost-effectiveness, acceptability to patients and clinicians and their impact on timely ovarian cancer detection and survival need to be evaluated. Currently, a large, clustered, randomised control trial is seeking to help to address this by investigating the clinical impact of implementing a suite of electronic cancer risk assessment tools (including an electronic version of the Hamilton ovarian SI) in UK general practice [45]. Studies have also sought to evaluate the impact of using tools as part of 'symptom-triggered screening' programs, but none have taken the form of randomised control trials—the gold standard approach—and so findings should

be interpreted with caution. In one study, 5000 women were approached in primary care clinics and screened for symptoms using the Goff SI, with further investigations performed if the Goff SI was positive [11]. However, conclusions were limited as only two ovarian cancers were identified in the study window. The Diagnosing Ovarian and Endometrial Cancer Early (DOvEE) trial also employs a proactive symptom-triggered testing approach, supported by media campaigns, in which women can self-refer and are screened for range of symptoms prior to study inclusion. Although the final DOvEE results are yet to be published, a pilot study reported that participants had lower tumour burden and more resectable disease than women diagnosed via the standard clinical pathway [9].

When considering the clinical utility of a tool, it is important to assess the proportion of women who are 'tool-positive' who ultimately have ovarian cancer, i.e., the PPV. Primary care cohort studies indicated that between 1 in 200 and 1 in 100 women who were QCancer tool-positive (5% or 10% risk) had the disease. Although these figures may appear low, evidence indicates that patients would opt for cancer testing at PPVs of 1% [46]. Further, having a positive tool result in the clinical setting does not necessarily mean that further investigation will automatically occur, as there may be a clear alternative cause for the symptoms—the tool is simply intended as a diagnostic aid to highlight the risk of ovarian cancer to the clinician. In addition, the most common follow-up tests—CA125 and transvaginal ultrasound—are relatively non-invasive, and CA125 is known to perform well when used in a symptomatic primary care population [47], although invasive investigations/surgery may ultimately be needed to determine whether ovarian cancer is present. In proactive symptom-triggered screening programs, the tool is more than just a diagnostic aid—it is the initial screening step which will dictate whether further ovarian cancer tests take place. The two population studies reporting PPVs relied on external ovarian cancer prevalence figures, but their PPV estimates were similar to that reported in the pilot DOvEE study (0.76% in women ≥ 50 years) [9]. Further research is needed to help determine whether, given this PPV, follow-up testing in proactive symptom-triggered testing programs is acceptable to women and improves outcomes. The definitive diagnosis of ovarian cancer often involves invasive procedures/surgery, which has contributed to patient morbidity in key ovarian cancer screening trials [3,39]. Although initial findings indicate that proactive symptom triggered testing approaches lead to minimal unnecessary surgery [9,11], large trials are needed to confirm that the implementation of symptom-based tools in clinical practice does not lead to significant excess morbidity.

6. Conclusions

Over 20 symptom-based tools have been developed in different populations to help assess women for ovarian cancer, but the majority have not been validated. Four symptom-based tools—the Goff SI, a modified version of the Goff symptom Index, SGO consensus criteria and QCancer Ovarian—have undergone independent external validation and exhibit similar sensitivities and specificities. These tools could have an important role to play in the detection of ovarian cancer, but further large well-conducted studies are needed to assess their cost-effectiveness, their acceptability, their effect on the timeliness of ovarian cancer diagnosis and their impact on clinical outcomes, including patient survival.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6694/12/12/3686/s1, Table S1: PRISMA Checklist. Text S1: MEDLINE search strategy. Table S2: Specific study exclusions. Table S3: Tool specifications. Table S4: Deviations from the original Goff SI in validation studies.

Author Contributions: Conceptualization, G.F.; methodology, G.F., F.M.W. and W.H.; formal analysis, G.F. and V.H.; data curation, G.F. and V.H.; writing—original draft preparation, G.F.; writing—review and editing, G.F., V.H., G.A., E.J.C., W.H., F.M.W. and J.E.; supervision, F.M.W., G.A., E.J.C. and W.H.; funding acquisition, F.M.W. and W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research arises from the CanTest Collaborative, which is funded by Cancer Research UK [C8640/A23385], of which G.F. is Clinical Research Fellow, V.H. is a PhD student, G.A. is the Senior Statistician, J.E. is Associate Director, and W.H. and F.M.W. are Directors. E.J.C. is supported through the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). The funders of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Conflicts of Interest: Two studies included in this review were conducted by W.H. W.H. played no role in study selection or quality assessment. All other authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Cancer Research UK. Ovarian Cancer Survival Statistics. Available online: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three (accessed on 20 May 2020).
3. Jacobs, I.J.; Menon, U.; Ryan, A.; Gentry-Maharaj, A.; Burnell, M.; Kalsi, J.K.; Amso, N.N.; Apostolidou, S.; Benjamin, E.; Cruickshank, D.; et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* **2016**, *387*, 945–956. [CrossRef]
4. Pinsky, P.F.; Yu, K.; Kramer, B.S.; Black, A.; Buys, S.S.; Partridge, E.; Gohagan, J.; Berg, C.D.; Prorok, P.C. Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. *Gynecol. Oncol.* **2016**, *143*, 270–275. [CrossRef] [PubMed]
5. Barrett, J.; Sharp, D.J.; Stapley, S.; Stabb, C.; Hamilton, W. Pathways to the diagnosis of ovarian cancer in the UK: A cohort study in primary care. *BJOG* **2010**, *117*, 610–614. [CrossRef]
6. National Cancer Intellegence Network. Routes to Diagnosis 2006–2016 by year, V2.1a. Available online: http://www.ncin.org.uk/publications/routes_to_diagnosis (accessed on 21 May 2020).
7. Funston, G.; Melle, V.M.; Ladegaard Baun, M.-L.; Jensen, H.; Helpser, C.; Emery, J.; Crosbie, E.; Thompson, M.; Hamilton, W.; Walter, F.M. Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: A systematic review of international guidelines. *BMC Cancer* **2019**, *19*, 1028. [CrossRef]
8. Ebell, M.H.; Culp, M.B.; Radke, T.J. A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer. *Am. J. Prev. Med.* **2016**, *50*, 384–394. [CrossRef]
9. Gilbert, L.; Basso, O.; Sampalis, J.; Karp, I.; Martins, C.; Feng, J.; Piedimonte, S.; Quintal, L.; Ramanakumar, A.V.; Takefman, J.; et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: Results from the prospective DOvE pilot project. *Lancet Oncol.* **2012**, *13*, 285–291. [CrossRef]
10. Goff, B.A.; Lowe, K.A.; Kane, J.C.; Robertson, M.D.; Gaul, M.A.; Andersen, M.R. Symptom triggered screening for ovarian cancer: A pilot study of feasibility and acceptability. *Gynecol. Oncol.* **2012**, *124*, 230–235. [CrossRef]
11. Andersen, M.R.; Lowe, K.A.; Goff, B.A. Value of Symptom-Triggered Diagnostic Evaluation for Ovarian Cancer. *Obstet. Gynecol.* **2014**, *123*, 73–79. [CrossRef]
12. Moons, K.G.M.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.A.; Macaskill, P.; Steyerberg, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [CrossRef]
13. Goff, B.A.; Mandel, L.S.; Melancon, C.H.; Muntz, H.G. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA* **2004**, *291*, 2705–2712. [CrossRef] [PubMed]
14. Hamilton, W.; Peters, T.J.; Bankhead, C.; Sharp, D. Risk of ovarian cancer in women with symptoms in primary care: Population based case-control study. *BMJ* **2009**, *339*, b2998. [CrossRef] [PubMed]
15. Laupacis, A.; Sekar, N.; Stiell, I.G. Clinical prediction rules: A review and suggested modifications of methodological standards. *J. Am. Med. Assoc.* **1997**, *277*, 488–494. [CrossRef]
16. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan-a web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [CrossRef]
17. Usher-Smith, J.A.; Sharp, S.J.; Griffin, S.J. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* **2016**, *353*, i3139. [CrossRef]
18. Whiting, P.F.; Rutjes, A.W.S.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.G.; Sterne, J.A.C.; Bossuyt, P.M.M. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [CrossRef]

19. Lurie, G.; Thompson, P.J.; McDuffie, K.E.; Carney, M.E.; Goodman, M.T. Prediagnostic symptoms of ovarian carcinoma: A case-control study. *Gynecol. Oncol.* **2009**, *114*, 231–236. [CrossRef]
20. Rossing, M.A.; Wicklund, K.G.; Cushing-Haugen, K.L.; Weiss, N.S. Predictive value of symptoms for early detection of ovarian cancer. *J. Natl. Cancer Inst.* **2010**, *102*, 222–229. [CrossRef]
21. Jordan, S.J.; Coory, M.D.; Webb, P.M. Re: Predictive Value of Symptoms for Early Detection of Ovarian Cancer. *JNCI J. Natl. Cancer Inst.* **2010**, *102*, 1599–1601. [CrossRef]
22. Hippisley-Cox, J.; Coupland, C. Identifying women with suspected ovarian cancer in primary care: Derivation and validation of algorithm. *BMJ* **2012**, *344*, d8009. [CrossRef]
23. Hippisley-Cox, J.; Coupland, C. Symptoms and risk factors to identify women with suspected cancer in primary care: Derivation and validation of an algorithm. *Br. J. Gen. Pract.* **2013**, *63*, e11–e21. [CrossRef]
24. Collins, G.S.; Altman, D.G. Identifying women with undetected ovarian cancer: Independent and external validation of QCancer® (Ovarian) prediction model. *Eur. J. Cancer Care (Engl.)* **2013**, *22*, 423–429. [CrossRef]
25. Grewal, K.; Hamilton, W.; Sharp, D. Ovarian cancer prediction: Development of a scoring system for primary care. *BJOG An Int. J. Obstet. Gynaecol.* **2013**, *120*, 1016–1019. [CrossRef]
26. Kim, M.-K.; Kim, K.; Kim, S.M.; Kim, J.W.; Park, N.-H.; Song, Y.-S.; Kang, S.-B. A hospital-based case-control study of identifying ovarian cancer using symptom index. *J. Gynecol. Oncol.* **2009**, *20*, 238–242. [CrossRef]
27. Macuks, R.; Baidekalna, I.; Donina, S. Diagnostic test for ovarian cancer composed of ovarian cancer symptom index, menopausal status and ovarian cancer antigen CA125. *Eur. J. Gynaecol. Oncol.* **2011**, *32*, 286–288.
28. Shetty, J.; Priyadarshini, P.; Pandey, D.; Manjunath, A.P. Modified Goff Symptom Index: Simple triage tool for ovarian malignancy. *Sultan Qaboos Univ. Med. J.* **2015**, *15*, e370–e375. [CrossRef]
29. Jain, S.; Danodia, K.; Suneja, A.; Mehndiratta, M.; Chawla, S. Symptom index for detection of ovarian malignancy in indian women: A hospital-based study. *J. Indian Acad. Clin. Med.* **2018**, *19*, 27–32.
30. Goff, B.A.; Mandel, L.S.; Drescher, C.W.; Urban, N.; Gough, S.; Schurman, K.M.; Patras, J.; Mahony, B.S.; Robyn Andersen, M. Development of an ovarian cancer symptom index: Possibilities for earlier detection. *Cancer* **2007**, *109*, 221–227. [CrossRef]
31. Andersen, M.R.; Goff, B.A.; Lowe, K.A.; Scholler, N.; Bergan, L.; Dresher, C.W.; Paley, P.; Urban, N. Combining a symptoms index with CA 125 to improve detection of ovarian cancer. *Cancer* **2008**, *113*, 484–489. [CrossRef]
32. Andersen, M.R.; Goff, B.A.; Lowe, K.A.; Scholler, N.; Bergan, L.; Drescher, C.W.; Paley, P.; Urban, N. Use of a Symptom Index, CA125, and HE4 to predict ovarian cancer. *Gynecol. Oncol.* **2010**, *116*, 378–383. [CrossRef] [PubMed]
33. Lim, A.W.W.; Mesher, D.; Gentry-Maharaj, A.; Balogun, N.; Jacobs, I.; Menon, U.; Sasieni, P. Predictive value of symptoms for ovarian cancer: Comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J. Natl. Cancer Inst.* **2012**, *104*, 114–124. [CrossRef] [PubMed]
34. Merritt, M.A.; Green, A.C.; Nagle, C.M.; Webb, P.M.; Bowtell, D.; Chenevix-Trench, G.; Green, A.; Webb, P.; DeFazio, A.; Gertig, D.; et al. Talcum powder, chronic pelvic inflammation and NSAIDs in relation to risk of epithelial ovarian cancer. *Int. J. Cancer* **2008**, *122*, 170–176. [CrossRef] [PubMed]
35. QResearch The QResearch Database. Available online: https://www.qresearch.org (accessed on 8 January 2019).
36. THIN® The Health Improvement Network. Available online: https://www.the-health-improvement-network.com/en/ (accessed on 9 November 2020).
37. Lowe, K.A.; Shah, C.; Wallace, E.; Anderson, G.; Paley, P.; McIntosh, M.; Andersen, M.R.; Scholler, N.; Bergan, L.; Thorpe, J.; et al. Effects of personal characteristics on serum CA125, mesothelin, and HE4 levels in healthy postmenopausal women at high-risk for ovarian cancer. *Cancer Epidemiol. Biomark. Prev.* **2008**, *17*, 2480–2487. [CrossRef]
38. Menon, U.; Gentry-Maharaj, A.; Hallett, R.; Ryan, A.; Burnell, M.; Sharma, A.; Lewis, S.; Davies, S.; Philpott, S.; Lopes, A.; et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: Results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol.* **2009**, *10*, 327–340. [CrossRef]
39. Buys, S.S.; Partridge, E.; Greene, M.H.; Prorok, P.C.; Reding, D.; Riley, T.L.; Hartge, P.; Fagerstrom, R.M.; Ragard, L.R.; Chia, D.; et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: Findings from the initial screen of a randomized trial. *Am. J. Obstet. Gynecol.* **2005**, *193*, 1630–1639. [CrossRef]

40. Jacobs, I.; Prys Davies, A.; Bridges, J.; Stabile, I.; Fay, T.; Lower, A.; Grudzinskas, J.G.; Oram, D. Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. *Br. Med. J.* **1993**, *306*, 1030–1034. [CrossRef]
41. *Cancer Incidence Projections Australia 2002 to 2011*; Australian Institute of Health and Welfare; Australasian Association of Cancer Registries; National Cancer Strategies Group: Canberra, Australia, 2005.
42. *Australian Demographic Statistics*; Australian Bureau of Statistics D: Canberra, Australia, 2007.
43. Chima, S.; Milley, K.; Reece, J.C.; Milton, S.; McIntosh, J.G.; Emery, J.D. Decision support tools to improve cancer diagnostic decision making in primary care: A systematic review. *Br. J. Gen. Pract.* **2019**, *69*, e809–e818. [CrossRef]
44. Chiang, P.C.; Glance, D.; Walker, J.; Walter, F.M.; Emery, J.D. Implementing a qcancer risk tool into general practice consultations: An exploratory study using simulated consultations with Australian general practitioners. *Br. J. Cancer* **2015**, *112*, S77–S83. [CrossRef]
45. Hamilton, W. Electronic Risk Assessment for Cancer for Patients in General Practice (ISRCTN22560297). Available online: http://www.isrctn.com/ISRCTN22560297 (accessed on 30 October 2020).
46. Banks, J.; Hollinghurst, S.; Bigwood, L.; Peters, T.J.; Walter, F.M.; Hamilton, W. Preferences for cancer investigation: A vignette-based study of primary-care attendees. *Lancet Oncol.* **2014**, *15*, 232–240. [CrossRef]
47. Funston, G.; Hamilton, W.; Abel, G.; Crosbie, E.J.; Rous, B.; Walter, F.M. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. *PLoS Med.* **2020**, *17*, e1003295. [CrossRef]

Garth Funston, Luke TA Mounce, Sarah Price, Brian Rous, Emma J Crosbie, Willie Hamilton and Fiona M Walter

CA125 test result, test-to-diagnosis interval, and stage in ovarian cancer at diagnosis:

a retrospective cohort study using electronic health records

Abstract

Background
In the UK, the cancer antigen 125 (CA125) test is recommended as a first-line investigation in women with symptoms of possible ovarian cancer.

Aim
To compare time between initial primary care CA125 test and diagnosis, tumour morphology, and stage in women with normal (<35 U/ml) and abnormal (≥35 U/ml) CA125 levels prior to ovarian cancer diagnosis.

Design and setting
Retrospective cohort study using English primary care and cancer registry data.

Method
Associations between CA125 test results and test-to-diagnosis interval, stage, and ovarian cancer morphology were examined.

Results
In total, 456 women were diagnosed with ovarian cancer in the 12 months after having a CA125 test. Of these, 351 (77%) had an abnormal, and 105 (23%) had a normal, CA125 test result. The median test-to-diagnosis interval was 35 days (interquartile range [IQR] 21–53) for those with abnormal CA125 levels, and 64 days (IQR 42–127) for normal CA125 levels. Tumour morphology differed by CA125 result: indolent borderline tumours were less common in those with abnormal CA125 levels (n = 47, 13%) than those with normal CA125 levels (n = 51, 49%) (P<0.001). Staging data were available for 304 women with abnormal, and 77 with normal, CA125 levels. Of those with abnormal CA125 levels, 35% (n = 106) were diagnosed at an early stage, compared to 86% (n = 66) of women with normal levels. The odds of being diagnosed with early-stage disease were higher in women with normal as opposed to abnormal CA125 levels (odds ratio 12.2, 95% confidence interval = 5.8 to 25.1, P<0.001).

Conclusion
Despite longer intervals between testing and diagnosis, women with normal, compared with abnormal, CA125 levels more frequently had indolent tumours and were more commonly diagnosed at an early stage in the course of the disease. Although testing approaches that have greater sensitivity might expedite diagnosis for some women, it is not known if this would translate to earlier-stage diagnosis.

Keywords
CA125; cancer antigen 125; diagnostic intervals; early diagnosis; general practice; ovarian cancer.

INTRODUCTION

Ovarian cancer is the sixth most common cancer to affect women in the UK, with >7000 women diagnosed each year.[1] It has the worst prognosis of all gynaecological cancers, accounting for >4000 UK deaths annually.[2] Although, overall, ovarian cancer prognosis is relatively poor, this varies markedly based on tumour type: studies conducted in the US and Sweden report that 5-year relative survival rates are 48% for invasive epithelial cancer (the most common type), compared with 93% for ovarian germ-cell tumours and 97% for borderline tumours.[3,4]

Most women with ovarian cancer are diagnosed after presenting with symptoms in primary care. However, the symptoms — such as bloating and abdominal pain — are non-specific and, therefore, have relatively low positive predictive values for the disease.[5,6] In 2011, the National Institute for Health and Care Excellence (NICE) advocated testing for the serum biomarker cancer antigen 125 (CA125) in women with symptoms of possible ovarian cancer in primary care.[7] NICE recommended that women with an elevated CA125 (≥35 U/ml) should undergo ultrasound testing;[7] however, they did not provide guidance on the follow-up or investigation of women with 'normal' (<35 U/ml) CA125 levels. Many other countries — including Ireland, Australia, Canada, and the US — also recommend CA125 as a primary care test for ovarian cancer.[8]

CA125 is a glycoprotein found in healthy ovaries, but blood levels commonly increase in ovarian cancer; around 80% of women with ovarian cancer have raised CA125 levels pre-surgery.[9] CA125 is more frequently elevated in advanced, rather than early-stage, disease and in some tumour types than others.[10] Concerns have been expressed that using CA125 as a single first-line investigation might delay diagnosis and lead to worse outcomes in women whose ovarian cancer is not associated with CA125 levels ≥35 U/ml,[11] yet there is little research exploring the relationship between CA125, time to diagnosis, and outcomes.

In this study, the authors examined the association of initial primary care pre-diagnostic CA125 results with the time between testing and diagnosis (test-to-diagnosis interval), tumour morphology, and disease stage in women with ovarian cancer.

METHOD

Study design, setting, and data sources

This retrospective cohort study utilised data from the Clinical Practice Research Datalink (CPRD) GOLD database, a dataset containing

G Funston, MB BChir, clinical research fellow; **FM Walter**, FRCGP, principal researcher in primary care cancer research, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge. **LTA Mounce**, PhD, research fellow; **S Price**, PhD, research fellow; **W Hamilton**, FRCP, FRCGP, professor of primary care diagnostics, University of Exeter Medical School, University of Exeter, Exeter. **B Rous**, FRCPath, consultant histopathologist, National Cancer Registration and Analysis Service, Public Health England, Cambridge. **EJ Crosbie**, FRCOG, professor of gynaecological oncology, Gynaecological Oncology Research Group, University of Manchester; consultant in gynaecological oncology, Manchester University NHS Foundation Trust, Manchester.

Address for correspondence
Garth Funston, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, 2 Worts' Causeway, Cambridge CB2 1TN, UK.
Email: gf272@medschl.cam.ac.uk

Submitted: 14 September 2020; **Editor's response:** 1 November 2020; **final acceptance:** 2 December 2020.

How this fits in

Cancer antigen 125 (CA125) is used as an initial test for women who present to primary care with symptoms of possible ovarian cancer, but research has shown that CA125 levels are normal in 23% of women prior to diagnosis. In the present study it was found that, although women with normal CA125 test results take longer to receive a diagnosis after testing than those with abnormal results, they are more likely to have less aggressive, more-curable forms of disease, and be diagnosised at an earlier cancer stage. This provides some reassurance for those using, and being tested for, CA125. However, improving the sensitivity of primary care testing approaches for ovarian cancer could still be of benefit to patients.

postcode-linked deprivation measures (provided by CPRD), and data from the National Cancer Registration and Analysis Service (NCRAS), which acts as the English cancer registry. CPRD GOLD comprises anonymised, coded primary care data, including laboratory results and diagnoses, for around 7% of the UK population.[12] The deprivation dataset consists of a five-level Townsend score — an area-level deprivation metric, in which higher scores indicate greater material deprivation. NCRAS data consists of detailed information on cancers diagnosed in England, including stage and morphology.[13] CPRD–NCRAS linkage was performed at patient level by NHS Digital.[14]

In order to match the coverage of NCRAS, this study was restricted to England.

Study period and cohort

A data sample obtained for a related study[15] was used. The sample consisted of women with a CA125 test recorded in CPRD GOLD between 1 May 2011 and 31 December 2014. From this sample, the following were excluded:

- women aged <18 years;
- those registered at a GP practice not deemed by the CPRD to be 'up to standard' regarding data quality;[12]
- those with a record of ovarian cancer on, or before, the CA125 test date; and
- women who had a CA125 test in the 12 months prior to the first CA125 test during the study period.

In order to maximise data quality, only CA125 entries recorded in standard CA125 units (U/ml, IU/ml, KU/L, or KIU/L) and with a laboratory upper reference limit were accepted. Similarly, CA125 values associated with clearly erroneous upper reference limits (such as 245 U/ml, 420 U/ml, and 455 U/ml) were excluded, as these could also indicate issues with the recording or coding of CA125 values.[15] The authors then identified women who had been diagnosed with ovarian cancer, as recorded in NCRAS data, within 12 months of CA125 testing. This group formed the study cohort.

Ovarian cancer, on the basis of codes from the tenth revision of the International Classification of Diseases (ICD-10), was defined as an ovarian malignancy (C56), a fallopian tube malignancy (C57.0), a peritoneal malignancy (C48.1 and C48.2), or a neoplasm of uncertain behaviour of the ovary (D39.1).[15] Fallopian and peritoneal cancers arise from the same tissue type and are diagnosed, staged, and treated in the same way as cancer arising from the surface of the ovary.

Borderline tumours are non-invasive, usually diagnosed at an early stage, and have a good prognosis. However, these may recur and, generally, require surgery. Borderline tumours are included in NICE guidance on ovarian cancer detection.[7]

CA125 category

NICE recommends using a CA125 cut-off of 35 U/ml.[7] Therefore, women were classified on the basis of the initial CA125 test into two groups:

- abnormal: CA125 level of ≥35 U/ml; and
- normal: CA125 level of <35 U/ml.

Covariates

A code list was used to identify symptoms of possible ovarian cancer included in current NICE guidelines[16] — namely, abdominal/pelvic pain, abdominal distension/bloating, change in bowel habit, fatigue, weight loss, urinary frequency/urgency, loss of appetite, pelvic mass, or ascites — that had been recorded in CPRD GOLD in the 30 days prior to CA125 testing. Level of deprivation was determined using the five-level Townsend score in the deprivation measures dataset.

Test-to-diagnosis interval

The date of cancer diagnosis is recorded for all tumours in NCRAS data. The test-to-diagnosis interval (days from first CA125 test in the year before diagnosis to diagnosis date, as recorded in NCRAS data) was calculated for all women.

Cancer stage and morphology

Tumour behaviour, morphology, and stage were identified from the NCRAS data.

Tumours were classified on the basis of ICD-10 codes as: 'borderline epithelial', 'invasive epithelial', 'invasive non-epithelial', and 'invasive not otherwise specified (NOS)'. Stage was categorised as early (stage I–II) or late (stage III–IV).

Statistical analysis

Accelerated failure time (AFT) models were used to examine the association between CA125 test results and test-to-diagnosis intervals. AFT models are a parametric time-to-event analysis previously utilised in CPRD research.[17] AFT models can be used to calculate time ratios. A time ratio >1 indicates that a variable prolongs the time to an event (for example, diagnosis), whereas a ratio <1 indicates that the variable reduces the time to the event. A univariate model was constructed to examine the relationship between the CA125 test result and test-to-diagnosis interval. A multivariable model was constructed incorporating age, a binary variable denoting the presence/absence of relevant symptoms prior to CA125 testing, and Townsend score. The presence or absence of a symptom was included as there is evidence that symptoms are more likely to be coded, rather than recorded in free text (which is unavailable for research), when they are more severe/persistent — which could result in expedited referral and diagnosis.[18] Weibull, generalised gamma, log-normal, and log-logistic distributions were examined. Log-logistic distribution was the best-fit parameterisation, according to the Akaike information criterion. Time ratios with associated *P*-values and 95% confidence intervals (CIs) were reported.

Fisher's exact test was used to assess whether women with abnormal and normal CA125 test results differed significantly in tumour morphology. Pairwise analyses were then performed to assess whether there was a statistically significant difference for each morphology category. The authors corrected for multiple comparisons, setting the significance level at $P = 0.01$.[19]

In a subgroup for whom stage data were recorded, logistic regression was used to examine the association between the CA125 test result and the disease stage at diagnosis. Adjustments were made for age, the presence/absence of a recorded symptom, and the Townsend score. Given the favourable prognosis of borderline tumours, a subanalysis was performed that excluded these. The authors explored the relationship between explanatory variables with missing stage data using logistic regression. Crude and adjusted odds ratios (ORs) with 95% CIs and associated *P*-values are reported.

All analyses were performed using Stata (version 15.1).

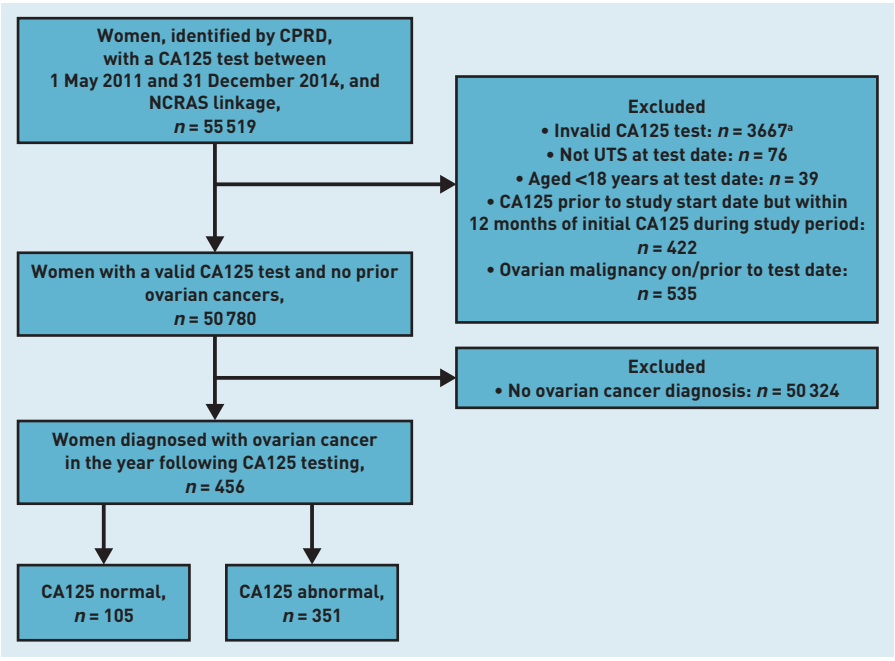


Figure 1. Application of selection criteria. [a]No CA125 value, no or incorrect units, or no or spurious upper threshold recorded. CA125 = cancer antigen 125. CPRD = Clinical Practice Research Datalink. NCRAS = National Cancer Registration and Analysis Service. UTS = up to standard.

RESULTS

The CPRD provided data on 55 519 women who were eligible for NCRAS linkage and who had a CA125 test between 1 May 2011 and 31 December 2014. After exclusions, 456 women diagnosed with ovarian cancer in the 12 months following CA125 testing were included in the study (Figure 1). Of these, 105 women (23%) had a normal initial CA125 result and 351 (77%) an abnormal CA125 result. A total of 41 (9%) women had a repeat CA125 test performed prior to diagnosis. Thirty women with an abnormal initial CA125 test result had a repeat test; for 29 (97%) of these, the result of the repeat test was also abnormal. Eleven women with a normal initial CA125 test had a repeat test and eight (73%) of these had an increase in their CA125 level; however, in only three cases (27%) was this increase sufficient to reach the ≥35 U/ml threshold (data not shown).

Mean age was higher in those with abnormal CA125 test results than those with normal CA125 test results, and a greater proportion of women with abnormal CA125 test results had a coded symptom of possible ovarian cancer (Table 1).

Test-to-diagnosis interval

The overall median test-to-diagnosis interval in the cohort was 42 days (interquartile range [IQR] 25–62) (data not shown). The interval was 35 days (IQR 21–53) for those with abnormal CA125 test results and 64 days

Table 1. Patient groups and baseline characteristics

CA125 test result	*n*	Mean age at diagnosis, years (range)	Patients with a symptom of possible ovarian cancer recorded pre-testing, *n* (%)[a]	Townsend score, *n* (%)				
				Level 1	Level 2	Level 3	Level 4	Level 5
Abnormal	351	65 (22–93)	212 (60)	80 (23)	100 (28)	78 (22)	61 (17)	32 (9)
Normal	105	57 (18–87)	59 (56)	24 (23)	31 (30)	25 (24)	14 (13)	11 (10)
Overall cohort	456	63 (18–93)	271 (59)	104 (23)	131 (29)	103 (23)	75 (16)	43 (9)

CA125 = cancer antigen 125.

Table 2. Median intervals by CA125 test result, and crude and adjusted associations between CA125 test result and test-to-diagnosis interval

CA125 test result	*n*	Median test-to-diagnosis interval in days, *n* (IQR)	Unadjusted association		Adjusted association[a]	
			Time ratio (95% CI)	*P*-value	Time ratio (95% CI)	*P*-value
Abnormal	351	35 (21–53)	Reference	<0.001	Reference	<0.001
Normal	105	64 (42–127)	2.0 (1.7 to 2.4)	—	2.0 (1.6 to 2.4)	—

[a]Adjusted for age, presence/absence of a recorded symptom, and Townsend score. Individual associations for all variables are displayed in Supplementary Table S1. CA125 = cancer antigen 125. IQR = interquartile range.

(IQR 42–127) for those with normal CA125 test results (Table 2). AFT models demonstrated a statistically significant association between CA125 test results and the test-to-diagnosis interval. A time ratio of 2.0 (95% CI = 1.7 to 2.4, *P*<0.001) indicated that the test-to-diagnosis interval for those women with normal CA125 test results was twice as long as for those with abnormal CA125 test results. The time ratio remained unaltered when adjusting for age, the presence/absence of a recorded symptom, and Townsend score.

Tumour morphology

Tumour morphology differed significantly, in statistical terms, by CA125 result (*P*<0.001) (Table 3). Invasive epithelial cancers were the most common type in women with abnormal CA125 test results (81%), whereas borderline tumours were the most common type in women with normal CA125 test results (49%). Serous tumours accounted for 52% of invasive tumours in those with an abnormal CA125 test result, compared with 30% in those with a normal CA125 test result (data not shown).

Stage at diagnosis

Staging information was missing for 75 women: 47 with an abnormal CA125 test result and 28 with a normal CA125 test result. In women with an abnormal CA125 test result in whom stage was recorded (*n* = 304), 106 (35%) were diagnosed with early-stage disease. In women with a normal CA125 test result in whom stage was recorded (*n* = 77), 66 (86%) were diagnosed with early-stage disease (data not shown).

Logistic regression, performed on data for patients with recorded disease stage and adjusted for age, the presence/absence of a recorded symptom, and Townsend score demonstrated that the odds of being diagnosed with early-stage disease were 12.2 times higher in women with normal than abnormal CA125 test results (Table 4). A subanalysis conducted after excluding borderline tumours demonstrated a statistically significant association between having a normal CA125 test result and being diagnosed at an early stage (OR 9.0, 95% CI = 4.0 to 19.8) (see Supplementary Table S2).

There was strong evidence to support an association between having a normal CA125 test result and having missing cancer stage at diagnosis in a logistic regression model; no such association was identified when borderline tumours were excluded from analysis (data not shown).

DISCUSSION

Summary

Women with normal CA125 test results in primary care, prior to receiving a diagnosis of ovarian cancer, took twice as long to be

Table 3. Tumour morphology by CA125 test result

	n	Borderline tumour, *n* (%)	Invasive tumour: Epithelial, *n* (%)	Invasive tumour: Non-epithelial, *n* (%)	Invasive tumour: NOS, *n* (%)	Overall analysis, *P*-value
Abnormal CA125 test result	351	47 (13)	284[a] (81)	4 (1)	16 (5)	<0.001
Normal CA125 test result	105	51 (49)	39[b] (37)	9 (9)	6 (6)	—
Pairwise analysis, *P*-value	—	<0.001	<0.001	<0.001	0.6	—

[a]Serous, n = 158; endometrioid, n = 16; mucinous, n = 14; clear cell, n = 14; other epithelial, n = 13; epithelial cancers of unknown morphology, n = 69. [b]Serous, n = 16; endometrioid, n = 4; mucinous, n = 8; clear cell, n = 3, other epithelial, n = 4; epithelial cancers of unknown morphology, n = 4. P-values are derived from Fisher's exact test for independence. CA125 = cancer antigen 125. NOS = not otherwise specified, that is, could not be classified as epithelial or non-epithelial based on the information in the cancer registry.

Table 4. The association between CA125 test results, age, and the presence/absence of a recorded symptom with early (stage I–II) diagnosis

		Unadjusted		Adjusted[a]	
Variable	***n***	**OR (95% CI)**	***P*-value**	**OR (95% CI)**	***P*-value**
Abnormal CA125 test result	304	Reference	—	Reference	—
Normal CA125 test result	77	11.2 (5.7 to 22.1)	<0.001	12.2 (5.8 to 25.5)	<0.001
Age	—	0.95 (0.93 to 0.96)	<0.001	0.94 (0.92 to 0.96)	<0.001
No symptom record	148	Reference	—	—	—
Symptom record	233	0.51 (0.33 to 0.77)	0.001	0.35 (0.21 to 0.59)	<0.001

[a]Model also adjusted for Townsend score. Data not shown for Townsend score as the variable was statistically insignificant (P = 0.9). CA125 = cancer antigen 125. OR = odds ratio.

diagnosed following testing as those with abnormal CA125 test results. Despite this, in women for whom staging data were available, 86% of those with normal CA125 test results were diagnosed at an early stage compared with only 35% of those with abnormal CA125 test results. In addition, indolent borderline ovarian tumours were more common, and aggressive invasive epithelial cancers were less common, in women with normal CA125 test results than in women with abnormal CA125 test results.

Strengths and limitations

A major strength of this study is its large size — the sample is equivalent to >6% of all ovarian cancers diagnosed in the UK each year. The results should be generalisable to women tested for CA125 in primary care prior to ovarian cancer diagnosis, as the primary care database used is generally representative of the UK population.[12] In addition, ovarian cancer diagnoses were identified from NCRAS, which reports a near-100% case ascertainment.[13]

This study does, however, have some limitations. When defining the cohort, it was assumed that cancer diagnosed within 12 months of the initial CA125 test was present at the time of testing. A period of 1 year, which has been used in similar studies[15,20,21] and was specified prior to data analysis,[15] was chosen as a compromise between minimising the inclusion of incidental cancers and maximising the inclusion of relevant cancers. Examining a longer follow-up was not possible as NCRAS data were only available until the end of 2015. However, given that only one woman out of 456 was diagnosed in month 12, extending follow-up is unlikely to alter the results. A shorter follow-up period — for example, 6 months — was not examined as this would have preferentially excluded patients from the group with normal CA125 test results (who have longer test-diagnosis intervals); the results would, therefore, have been biased.

Patients with severe disease, who often have severe symptoms, frequently experience expedited diagnoses when compared with those with less severe disease — an observation sometimes referred to as the 'sick quick' phenomenon.[22] As CA125 levels are also more likely to be elevated in women with more-severe disease, this may act as a confounder. The analyses were adjusted for the presence/absence of relevant coded symptoms, as symptoms may be more likely to be coded (rather than mentioned in free text) if they are more severe,[18] but it is unlikely that it was possible to adjust fully for severity of symptoms and disease.

The authors considered adjusting for ethnicity in the analyses, but this was not done as not all patients have an ethnicity recorded in CPRD GOLD.[23] The authors are not aware of any evidence within the literature indicating that ethnicity is associated with either diagnostic interval or stage at diagnosis for ovarian cancer, and would not expect the inclusion of ethnicity to markedly alter the results.

There was a statistically significant association between having a normal CA125 result and having missing stage at diagnosis. This is to be expected, as stage is less frequently recorded in the cancer registry for borderline tumours, which are more common in women with normal CA125 test results. It is reassuring that when borderline tumours were excluded no statistically significant association

between the CA125 result and missing stage was identified, and a normal CA125 result was still strongly associated with early-stage diagnosis. Although there is no reason to suspect that study findings would differ markedly if staging data were available for all patients, the magnitude of the association between CA125 result and stage should be interpreted with caution.

Comparison with existing literature

Previous research has identified an association between false negative results and longer healthcare intervals: in one study, patients with a negative chest X-ray who went on to be diagnosed with lung cancer experienced longer primary care intervals than those with an abnormal chest X-ray;[24] in another study, patients with a false negative rheumatoid factor in primary care, prior to a rheumatoid arthritis diagnosis, took longer to be referred to a specialist.[25] Research indicates that receiving an 'all clear' diagnosis (no cancer) following testing in primary care can provide reassurance to patients, which can lead to delayed re-presentation if symptoms persist or recur.[26] Similarly, false reassurance could affect GPs, prompting them to seek alternative diagnoses and delaying referral.[24,27]

Few studies have investigated the relationship between false negative results and cancer outcomes in patients who are symptomatic, although one — by Yeh *et al*[28] — did find that patients with false negative fine-needle aspiration results, who were then diagnosed with thyroid cancer, were more likely to have vascular and capsular invasion and experience persistent disease post-treatment.

In the study presented here, for the majority of women with normal CA125 test results, cancer was detected at an early stage; this was in contrast with women with abnormal results, despite those with normal CA125 test results having longer test-to-diagnosis intervals. This finding could be due to differences in tumour type. In the study presented here, borderline tumours were nearly four times as common in women with normal, rather than abnormal, levels of CA125. Borderline tumours less frequently cause elevations in CA125 than their invasive counterparts, tend to grow slowly, and 80% are diagnosed at an early stage.[29] In contrast, invasive epithelial tumours, which typically have an insidious onset and poor survival, were twice as common in women with abnormal than normal CA125 levels. Further, aggressive invasive serous tumours, which are more frequently diagnosed at a later stage and more frequently elevate CA125 levels than other invasive tumour types,[10] accounted for half of invasive tumours in women with abnormal CA125 test results and only a third of invasive cancers in women with normal CA125 test results.

The authors employed the NICE advocated threshold of 35 U/ml in the present study to categorise results as 'normal' or 'abnormal'. However, this is an oversimplification: recent research has shown that the probability of ovarian cancer is much higher in women with a CA125 level of 34 U/ ml compared with those with a CA125 level of 1 U/ml,[15] yet these results are all classified as 'normal' under NICE guidelines (and within this study). Newly developed CA125-based primary care prediction models could help select women for further investigation or referral (instead of the 35 U/ml threshold),[15] but require further evaluation.

Implications for research and practice

CA125 test results detected 77% of ovarian cancer cases in the cohort and 88% of the invasive epithelial subtype, which is responsible for the majority of ovarian cancer mortality.[30] Abnormal CA125 test results are, therefore, helpful in identifying women with possible ovarian cancer, especially the most lethal type. However, a normal CA125 test result does not exclude disease.

It is reassuring that most women with normal CA125 test results were diagnosed at an early stage, despite taking longer to be diagnosed. However, given the observational nature of this study, it was not possible to determine to what extent women with normal CA125 test results experienced disease progression or worse survival rates as a result of their prolonged test-to-diagnosis intervals. Diagnostic strategies that use novel serum biomarkers or imaging modalities in combination with CA125 may detect additional ovarian cancer cases,[8,31] which could expedite diagnosis in some women. However, large, prospective studies would need to be undertaken to determine whether implementing more sensitive testing strategies would lead to earlier stage diagnosis and improved survival.

Regardless of its impact on survival, reducing unnecessary delay in ovarian cancer diagnosis is likely to be beneficial for women with normal CA125 test results. Delay in cancer diagnosis is associated with psychological distress, particularly among women,[32] and perceived delays can damage doctor–patient relationships.[33] Earlier diagnosis of ovarian cancer could

reduce morbidity, even if a stage shift is not achieved, by detecting lower-volume disease.[31] Possible strategies to reduce diagnostic delay could include appropriate safety netting with reassessment, and undertaking re-testing or alternative investigations (for example, ultrasound), if symptoms persist or worsen. In the study presented here, only a small proportion of women with normal results in their initial CA125 test had a repeat test. In 73% of repeat tests, there was an increase in CA125 levels, but in only 27% was this increase sufficient to reach the 35 U/ml threshold — this supports the idea that rising levels below the 35 U/ml threshold could be used to prompt further investigation.[31] The nature, duration, and severity of presentation should also be considered when deciding on a follow-up strategy. For example, if a patient develops a pelvic mass (which has a high positive predictive value for ovarian cancer) an urgent referral is warranted,[16,34] whereas alternative follow-up strategies, such as referral for ultrasound or CA125 re-testing, may be more appropriate for less highly predictive presentations.

Funding

This research arises from the CanTest Collaborative, which is funded by Cancer Research UK (reference number: C8640/A23385); Garth Funston is clinical research fellow, and Willie Hamilton and Fiona M Walter are directors of CanTest. This study was also funded by the National Institute for Health Research (NIHR) School of Primary Care Research (reference number: FR17 424). Emma J Crosbie is supported by the NIHR Manchester Biomedical Research Centre (reference number: IS-BRC-1215-20007). The funders of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or Cancer Research UK.

Ethical approval

Approval for this research was obtained from the Independent Scientific Advisory Committee for the Medicines and Healthcare products Regulatory Agency (protocol number: 18_184).

Provenance

Freely submitted; externally peer reviewed.

Competing interests

The authors have declared no competing interests.

Acknowledgements

The authors would like to acknowledge that this work uses data provided by patients and collected by the NHS as part of their care and support.

Discuss this article

Contribute and read comments about this article: **bjgp.org/letters**

REFERENCES

1. Cancer Research UK. Ovarian cancer incidence statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence (accessed 30 Mar 2021).
2. Cancer Research UK. Ovarian cancer mortality statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/mortality (accessed 30 Mar 2021).
3. American Cancer Society. Survival rates for ovarian cancer. 2021. https://www.cancer.org/cancer/ovarian-cancer/detection-diagnosis-staging/survival-rates.html (accessed 30 Mar 2021).
4. Kalapotharakos G, Högberg T, Bergfeldt K, Borgfeldt C. Long-term survival in women with borderline ovarian tumors: a population-based survey of borderline ovarian tumors in Sweden 1960–2007. *Acta Obstet Gynecol Scand* 2016; **95(4):** 473–479.
5. Public Health England. Routes to diagnosis 2006-2016 by year workbook. 2019. http://www.ncin.org.uk/publications/routes_to_diagnosis (accessed 30 Mar 2021).
6. Funston G, O'Flynn H, Ryan NAJ, *et al.* Recognizing gynecological cancer in primary care: risk factors, red flags, and referrals. *Adv Ther* 2018; **35(4):** 577–589.
7. National Institute for Health and Care Excellence. *Ovarian cancer: recognition and initial management. CG122.* 2011. https://www.nice.org.uk/guidance/cg122 (accessed 8 Apr 2021).
8. Funston G, Melle VM, Ladegaard Baun M-L, *et al.* Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. *BMC Cancer* 2019; **19(1):** 1028.
9. Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. Accuracy of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. *Eur J Obstet Gynecol Reprod Biol* 2009; **142(2):** 99–105.
10. Jacobs I, Bast Jr RC. The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 1989; **4(1):** 1–12.
11. Olaitan A. Rapid response to: Recognition and initial management of ovarian cancer: summary of NICE guidance. https://www.bmj.com/rapid-response/2011/11/03/recognition-and-initial-management-ovarian-cancer-summary-nice-guidance (accessed 8 Apr 2021).
12. Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44(3):** 827–836.
13. Henson KE, Elliss-Brookes L, Coupland VH, *et al.* Data resource profile: National Cancer Registration Dataset in England. *Int J Epidemiol* 2020; **49(1):** 16–16h.
14. Padmanabhan S, Carty L, Cameron E, *et al.* Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol* 2019; **34(1):** 91–99.
15. Funston G, Hamilton W, Abel G, *et al.* The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. *PLoS Med* 2020; **17(10):** e1003295.
16. National Institute for Health and Care Excellence. *Suspected cancer: recognition and referral. NG12.* 2021. https://www.nice.org.uk/guidance/ng12 (accessed 30 Mar 2021).
17. Ridd MJ, Santos Ferreira DL, Montgomery AA, *et al.* Patient–doctor continuity and diagnosis of cancer: electronic medical records study in general practice. *Br J Gen Pract* 2015; DOI: https://doi.org/10.3399/bjgp15X684829.
18. Price SJ, Stapley SA, Shephard E, *et al.* Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case-control study. *BMJ Open* 2016; **6(5):** e011664.
19. McDonald JH. *Handbook of biological statistics.* 3rd edn. Baltimore, MD: Sparky House Publishing, 2014.
20. Watson J, Salisbury C, Banks J, *et al.* Predictive value of inflammatory markers for cancer diagnosis in primary care: a prospective cohort study using electronic health records. *Br J Cancer* 2019; **120(11):** 1045–1051.
21. Neal RD, Din NU, Hamilton W, *et al.* Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer* 2014; **110(3):** 584–592.
22. Tørring ML, Frydenberg M, Hansen RP, *et al.* Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *Eur J Cancer* 2013; **49(9):** 2187–2198.
23. Mathur R, Bhaskaran K, Chaturvedi N, *et al.* Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf)* 2014; **36(4):** 684–692.
24. Bjerager M, Palshof T, Dahl R, *et al.* Delay in diagnosis of lung cancer in general practice. *Br J Gen Pract* 2006; **56(532):** 863–868.
25. Miller A, Nightingale AL, Sammon CJ, *et al.* Negative rheumatoid factor in primary care delays referral of patients with rheumatoid arthritis. *Rheumatology* 2014; **53(Suppl 1):** i29.
26. Renzi C, Whitaker KL, Winstanley K, *et al.* Unintended consequences of an 'all-clear' diagnosis for potential cancer symptoms: a nested qualitative interview study with primary care patients. *Br J Gen Pract* 2016; DOI: https://doi.org/10.3399/bjgp16X683845.
27. Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K. False-negative results in screening programmes: systematic review of impact and implications. *Health Technol Assess* 2000; **4(5):** 1–120.
28. Yeh MW, Demircan O, Ituarte P, Clark OH. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. *Thyroid* 2004; **14(3):** 207–215.
29. Fischerova D, Zikan M, Dundr P, Cibula D. Diagnosis, treatment, and follow-up of borderline ovarian tumors. *Oncologist* 2012; **17(12):** 1515–1533.
30. Elias KM, Guo J, Bast Jr RC. Early detection of ovarian cancer. *Hematol Oncol Clin North Am* 2018; **32(6):** 903–914.
31. Gilbert L, Basso O, Sampalis J, *et al.* Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *Lancet Oncol* 2012; **13(3):** 285–291.
32. Risberg T, Sørbye SW, Norum J, Wist EA. Diagnostic delay causes more psychological distress in female than in male cancer patients. *Anticancer Res* 1996; **16(2):** 995–999.
33. Norman A, Sisler J, Hack T, Harlos M. Family physicians and cancer care. Palliative care patients' perspective. *Can Fam Physician* 2001; **47:** 2009–2012, 2015–2016.
34. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based casecontrol study. *BMJ* 2009; **339:** b2998.

Could ovarian cancer prediction models improve the triage of symptomatic women in primary care? A modelling study using routinely collected data

Garth Funston[1*], Gary Abel[2], Emma J. Crosbie[3,4], Willie Hamilton[2], Fiona M. Walter[1]

[1]The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[2]University of Exeter, Exeter, UK
[3]Gynaecological Oncology Research Group, Division of Cancer Sciences, University of Manchester, Manchester, UK
[4]Department of Obstetrics and Gynaecology, Manchester University NHS Foundation Trust, Manchester Academic Health Sciences Centre, Oxford Road, Manchester, UK

*gf272@medschl.cam.ac.uk

Key words: General practice; obstetrics and gynaecology; risk management; decision support, clinical; women's health

ABSTRACT

Background

CA125 is recommended as the first line test in women presenting with symptoms of possible ovarian cancer in England. This study sought to develop and internally validate CA125 based diagnostic prediction models, and to explore potential diagnostic implications of implementing model-based thresholds for further investigation in primary care.

Methods

This retrospective cohort study used routinely collected primary care and cancer registry data from symptomatic, CA125 tested women in England (2011-2014). A total of 29,962 women were included, of whom 279 were diagnosed with ovarian cancer. Logistic regression was used to develop two models to estimate the probability of ovarian cancer: Model 1 consisted of age and CA125 level alone; Model 2 incorporated further risk factors. Model discrimination (AUC) was evaluated using 10-fold cross-validation. The sensitivity and specificity of various model ovarian cancer risk thresholds (≥1% - ≥3%) were compared with that of the current CA125 cut-off (≥35U/ml). The implications of applying these thresholds to our cohort were examined.

Results

Model 1 exhibited excellent discrimination (AUC: 0.94) on cross-validation. The inclusion of additional variables (Model 2) did not improve performance. At a risk threshold of ≥1% Model 1 exhibited greater sensitivity (86.4% vs 78.5%) but lower specificity (89.1% vs 94.5%) than CA125 (≥35U/ml). Applying the ≥1% model threshold to our cohort, in place of the current CA125 cut-off, 1 in every 74 additional women identified for further investigation had ovarian cancer.

Interpretation

A model incorporating age and CA125 level alone performed well for the identification of ovarian cancer in symptomatic women. Following external validation, it could be used as part of a 'risk-based triage' system, in which women at high risk of undiagnosed ovarian cancer are selected for urgent specialist investigation while women at 'low risk but not no

risk' are offered non-urgent investigation or interval CA125 re-testing. Such an approach has the potential to expedite ovarian cancer diagnosis, but further research is needed to evaluate the clinical impact and health-economic implications.

INTRODUCTION

Ovarian cancer is the 6th most common cancer to affect UK women and has the worst prognosis of any gynaecological malignancy.[1] Survival is highly dependent on the stage at diagnosis with five-year survivals of 90% and 4% for UK women diagnosed at stage I and IV respectively.[2] Although several large screening trials have been conducted, as yet they have failed to demonstrate a long term survival benefit.[3,4] In the absence of screening programs, the majority of women with ovarian cancer are diagnosed after they present to their GP with symptoms;[5,6] thus timely diagnosis of these women may improve cancer outcomes.

In 2011, the National Institute for Health and Care Excellence (NICE) recommended that women presenting to their GP with symptoms of possible ovarian cancer in England, Wales and Northern Ireland be tested for the serum biomarker Cancer Antigen 125 (CA125).[7] Further investigation with ultrasound was advocated if CA125 levels were ≥35 U/ml. However, this threshold was not based on primary care evidence. 23% of women with ovarian cancer have levels of less than 35 U/ml in primary care prior to diagnosis,[8] and safety concerns have been raised over the potential clinical impact of delayed diagnoses in this false negative group.[9] In addition, even if woman have markedly elevated CA125 levels (indicating a very high likelihood of undiagnosed ovarian cancer), they still must undergo GP-requested ultrasound before they are eligible for an urgent ('two week wait') cancer pathway referral to a specialist.

We recently developed a diagnostic prediction model based on age and CA125 level in order to estimate the probability of ovarian cancer in individual women undergoing testing in primary care in England.[8] If implemented in clinical practice, such models would allow women to be triaged using the risk of undiagnosed ovarian cancer, so that those at greatest risk undergo urgent specialist investigation, those at 'low risk but not no risk' are monitored or undergo non-urgent investigation, and those at very low risk (the majority) can be

reassured. Such primary care evidenced risk-based triage strategies may reduce diagnostic delay and thereby improve patient outcomes.

This study had three objectives. First, as previous studies have shown that a range of patient risk factors and blood tests can help predict undiagnosed ovarian cancer,[10] we sought to develop a comprehensive diagnostic prediction model, making best use of variables routinely available in primary care. Second, to compare the diagnostic performance of this model with that of a simpler model comprising age and CA125. Third, to explore the potential implications of implementing different model risk thresholds on ovarian cancer detection in primary care.

METHODS

This study is reported in accordance with transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines. A completed checklist is included (online supplementary appendix 1).

Data source

We conducted a retrospective cohort study using linked data from the Clinical Practice Research Datalink (CPRD) GOLD, Hospital Episode Statistics Admitted Patient Care dataset (HES APC) and the National Cancer Registration and Analysis Service (NCRAS). The CPRD GOLD dataset contains anonymised, coded, primary care data including demographics, laboratory results and symptoms for around 11 million patients and is broadly representative of the UK population.[11] The HES APC dataset includes information about hospital admissions and patient demographics, including ethnicity.[12] The NCRAS (the English cancer registry) collects cancer registration data on patients, including detailed information on tumour topography and diagnosis date. The NCRAS obtains data from hospitals, GP surgeries, and death certificates and reports a near 100% case ascertainment.[13] Linkage of CPRD and NCRAS data was performed at a patient level by a third party, National Health Service (NHS) Digital.[14] NCRAS only collects details of cancers diagnosed in England, so the study was restricted to English general practices.

Participants

We identified women with a code for CA125 measurement in primary care between 1st May 2011 and the 31st December 2014 and included all women who met study criteria in order to maximise sample size. We excluded women who were <18 years old or registered at a GP practice not deemed 'up-to-standard' on data quality by CPRD on the date of their first CA125 test during this period.[11] Women with a record of ovarian cancer in NCRAS data on or before the CA125 test date were also excluded, as were women with a CA125 test in the 12 months before the first CA125 test during the study period. To maximise data quality, only CA125 entries recorded in standard equivalent units of CA125 measurement (U/ml, IU/ml, KU/L, KIU/L) were accepted, and CA125 values associated with spurious laboratory cut-offs (245, 420 and 455 U/ml), or those where no cut-off was given, were excluded. As our aim was to develop models for use in symptomatic women, and we wished to develop a model which included symptom variables, we restricted the cohort to those who had a recognised symptom of possible ovarian cancer recorded in the year prior to CA125 testing. A code list was used to identify symptoms included in current NICE guidelines on ovarian cancer detection.[15] Women with ascites or a palpable pelvic mass recorded prior to CA125 testing were excluded, as these clinical signs have high positive predictive values for ovarian cancer and warrant urgent referral to a specialist.[15,16] Where women had more than one CA125 test during the study period, only the first was used in the analyses.

Outcome definition

The primary outcome was a diagnosis of ovarian cancer, as recorded using International Classification of Disease (ICD)-10 codes in NCRAS data, in the 12 months following initial CA125 testing. Ovarian cancer was defined as an ovarian malignancy (C56), a fallopian tube malignancy (C57.0), a peritoneal malignancy (C48.1, C48.2) or a neoplasm of uncertain behaviour of the ovary (D39.1).[17,18]

Prediction models

Candidate variables

From the literature, we identified a list of candidate variables which included: 1) established ovarian cancer risk/protective factors, 2) ovarian cancer symptoms and 3) blood tests. We focussed on variables which are well recorded within GP records. Routine blood tests were

included as recent studies have shown that both having a test performed and the specific test level can be predictive of undiagnosed cancer in primary care.[19–21] The final list of candidate variables taken forward into data-driven selection procedures was determined by the consensus of a multidisciplinary group consisting of GPs (GF, FMW, WH), a gynaecological oncologist (EJC) and a statistician (GA) (**Table 1**). Ethnicity was classified into 5 categories (White, Mixed, Black, Asian, Other) in line with the 2001 census, using a pre-developed code list.[22] These were subsequently collapsed into 2 groups - "White" and "Other ethnicities" - as numbers in individual ethnic groups, other than White, were small and the risk of ovarian cancer is higher in people of White ethnicity than Black and Asian ethnicity.[23] Recent research indicates that patients who have had routine blood tests performed in primary care are at greater risk of undiagnosed cancer (even if the test results are normal) when compared to those who have not had routine blood tests performed.[21] Therefore, we included each routine blood test as a categorical variable with "no test" forming a category. Full details of variable preparation are included in the supplemental material (online supplemental appendix 2).

Table 1. Candidate variables.

Variable	Data source	Categorisation	Variable inclusion time/period
Risk / protective factors			
Age[24]	CPRD	Continuous (years)	On date of CA125 testing
Ethnicity[23,25]	CPRD and HES	Categorical: *White* *Other ethnicities*	Most frequently recorded[22]
Height[26–29]	CPRD	Continuous (cm)	Most recent on/prior to CA125 test date recorded when ≥18years old
BMI[26,30,31]	CPRD	Continuous (Kg/m^2)	Most recent on/in the 10 years prior to CA125 test date ≥18 years old
Personal history breast cancer[32]	CPRD / NCRAS	Binary	Up to CA125 test date
Symptoms			
Ovarian cancer symptoms[7]	CPRD	Binary for each symptom. Presence/absence of: abdominal/pelvic pain, appetite loss, bloating, distension, change in bowel habit, fatigue, urinary frequency/urgency, new irritable bowel syndrome (≥50 years old), weight loss	12 months prior to CA125 testing
Blood biomarkers			
CA125[8]	CPRD	Continuous	First valid CA125 level in study period
Albumin[33]	CPRD	Categorical: *Not tested* *<35 g/L* *≥35 g/L*	Most recent record on or in the 12 months prior to the CA125 test date
Haemoglobin[34]	CPRD	Categorical: *Not tested* *<12g/dl* *≥12g/d*	Most recent record on or in the 12 months prior to the CA125 test date
Platelets[20]	CPRD	Categorical: *Not tested* *<300x10^9/L* *300-449x10^9/L* *≥450x10^9/L*	Most recent record on or in the 12 months prior to the CA125 test date
CRP[21,35]	CPRD	Categorical: *Not tested* *<3 mg/L* *3-9.99 mg/L* *≥10 mg/L*	Most recent record on or in the 12 months prior to the CA125 test date

Abbreviations: BMI = body mass index, CRP = C-reactive protein, CPRD=Clinical Practice Research Datalink, CA125= cancer antigen 125, NCAS= National Cancer Registration and Analysis Service. Included citations provide further details on the association between candidate variables and cancer risk. The choice of categories for categorical blood tests was based on standard references ranges and the literature.

Model derivation and internal validation

We developed two models. Model 1 was pre-specified and contained age and CA125 level alone. Model 2 contained the most predictive variables selected from the list of candidates.

Prior to model derivation, continuous variables were mean centred. BMI and CA125 level were right skewed, so were log transformed. The relationships between log CA125 and age with ovarian cancer were non-linear. To account for this, we generated restricted cubic splines (5-knots) for log CA125 and for age and included these in place of age and log CA125 within the models.[36]

In order to derive Model 2, multivariate imputation by chained equation (MICE)[37] was used to replace missing data on ethnicity, height and Body Mass Index (BMI). 20 imputations were performed. Following imputation, a logistic regression model, containing all candidate variables, was fitted to estimate variable coefficients. Rubin's rules were used to combine the results across the imputed datasets.[38] Possible interaction between age and log CA125 level was examined through the inclusion of an interaction term. We examined 19 candidate variables with a total of 32 degrees of freedom (main effect and non-main effect), giving 9 'Events Per Parameter' (EPP). To select variables for Model 2, a backward elimination approach was used in which the full model was fitted, the least significant variable was removed, then the model was refitted and the process repeated until all variables had a p value of ≤0.05. Variable coefficients were used as model weights.

To assess model discrimination i.e. the ability of a model to distinguish those who have a disease from those who do not have a disease, we calculated area under the receiver operating characteristic curve (AUC). 10-fold cross-validation was performed to assess for any optimism in model discrimination. To examine model calibration i.e. agreement between model estimated outcomes and observed outcomes, we calculated the calibration slope using 10-fold cross-validation. For Model 2, discrimination and calibration were calculated for each imputed dataset and Rubin's rules were used to combine results across the imputed datasets.

Model thresholds

We devised thresholds to identify women with a ≥1%, ≥2% and ≥3% probability of undiagnosed ovarian cancer, based on our models. The lowest threshold - ≥1% - was selected as patients have reported that they would opt for cancer investigations at this risk level.[39] The highest threshold - ≥3% - was chosen to match the 'risk threshold' at which NICE advocate urgent specialist cancer investigation or referral for symptomatic primary care patients.[15] We applied these thresholds to the study cohort to calculate their diagnostic accuracy (sensitivity, specificity, positive predictive value [PPV] and negative predictive value [NPV]) for the detection of ovarian cancer within 12 months of CA125 testing. We compared these accuracy metrics with that of CA125 at its standard, NICE advocated, cut-off (≥35U/ml).[7] We also compared the accuracy of model thresholds to that of CA125 cut-offs with equivalent sensitivities, to determine whether using the models, rather than just CA125 based cut-offs, improved accuracy.

All analysis was performed in Stata version 15.1. The user written *cvAUROC* command was used to calculate cross-validation AUCs.[40]

RESULTS

29,962 women met inclusion criteria (**Figure 1**). Of these, 279 (0.9%) were diagnosed with ovarian cancer in the 12 months following CA125 testing.

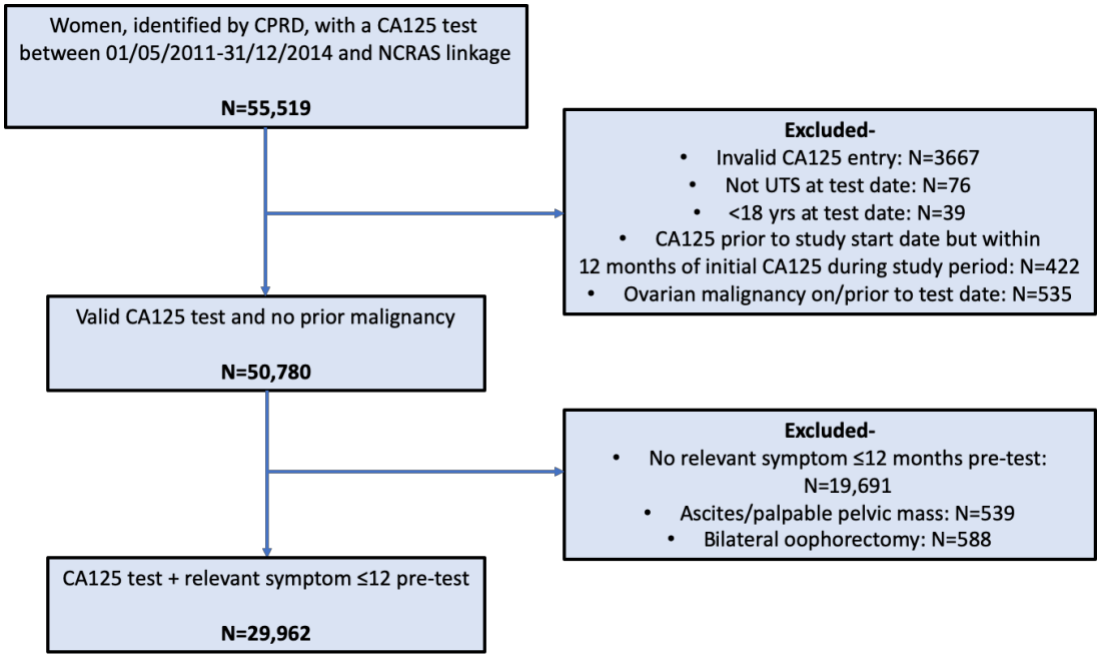


Figure 1. Application of study criteria.
"Invalid CA125 entry" = no CA125 value, no or incorrect units, no upper threshold or spurious upper threshold.

The baseline characteristics of the cohort are shown in **Table 2**. Missing data was noted for ethnicity (n=1,234, 4.1%), height (n=1,721, 5.7%) and BMI (n=2,986, 10%). The mean patient age was 55 years (range 18-101 years). The most common symptom was abdominal/pelvic pain, recorded for 58.5% of women. The proportion of women who had blood tests recorded (other than CA125) ranged from 66% for CRP to 94.5% for haemoglobin.

Table 2. Cohort baseline characteristics.

Variable	N=29,962
Risk / protective factors	
Age (years)	Mean=55 (SD:15)
Ethnicity:	
White	n=26,511 (88.5%)
*Other ethnicities**	n=2,217 (7.4%)
Height (cm)	Mean=162 (SD:6.8)
BMI (Kg/m^2)	Median=25.8 (IQR:22.8-29.7)
Personal history breast cancer	n=1,168 (3.9%)
Symptoms	
Abdominal / pelvic pain	n=17,538 (58.5%)
Appetite loss	n=203 (0.7%)
Bloating	n=5,649 (18.9%)
Distension	n=821 (2.7%)
CIBH	n=5,808 (19.4%)
Fatigue	n=3,968 (13.2%)
Urinary frequency / urgency	n=1,503 (5%)
≥50 years of age with new IBS	n=286 (1%)
Weight loss	n=1,168 (3.9%)
Blood biomarkers	
CA125	Median=12 (IQR:8-17)
Albumin:	
Not tested	n=3,723 (12.4%)
<35 g/L	n=834 (2.8%)
≥35 g/L	n=25,405 (84.8%)
Haemoglobin:	
Not tested	n=1,648 (5.5%)
<12g/dl	n=3,089 (10.3%)
≥12g/dl	n=25,225 (84.2%)
Platelets:	
Not tested	n=1,679 (5.6%)
<300x10^9/L	n=20,442 (68.2%)
300-449x10^9/L	n=7,314 (24.4%)
≥450x10^9/L	n=527 (1.8%)
CRP:	
Not tested	n=13,181 (44%)
<3 mg/L	n=6,907 (23.1%)
3-9.99 mg/L	n=7,370 (24.6%)
≥10 mg/L	n=2,504 (8.4%)

*Asian (n=1,117), Black (n=562), Mixed (n=160) and Other (n=378)
Abbreviations: SD= standard deviation, IQR= interquartile range, IBS= irritable bowel syndrome, CA125= cancer antigen 125, CRP= C-reactive protein.

Predictor variables

The variables retained in Model 2 after backward elimination procedures were: age, ethnicity, BMI, height, abdominal/pelvic pain, distension, CA125 level, platelet level and

albumin level. Coefficients and odds ratios for all variables are included in supplemental information (online supplemental appendix 3).

Model discrimination and calibration

The AUCs of Model 1 and 2, when directly calculated from the dataset (apparent performance), were similar (**Table 3**). On cross-validation the models had the same AUC (0.935). There was little difference between apparent and cross-validation AUCs, indicating that model overfitting/optimism was minimal. The AUC of CA125 alone, calculated directly from the study cohort, was 0.932. Models 1 and 2 had calibration slopes close to 1, indicating good calibration, but confidence intervals were wide.

Table 3. Model discrimination and calibration.

Model	Apparent AUC	Cross-validation AUC*	Cross-validation calibration slope (95% CI)
Model 1	0.938	0.935	1.01 (0.606-1.42)
Model 2	0.938	0.935	1.05 (0.673-1.42)

*Cross-validation ROC curves are included in online supplemental appendix 4.

Thresholds for further investigation

As the more parsimonious Model 1 exhibited the same AUC and similar calibration metrics to Model 2, the evaluation of thresholds for further investigation focussed on Model 1. The diagnostic accuracies of Model 1 thresholds, for the detection of ovarian cancer within 12 months of CA125 testing, are shown in **Table 4**. These were compared against CA125 cut-offs with equivalent sensitives for ovarian cancer. At the ≥1% probability threshold, the specificity of Model 1 was 3.1% higher than a CA125 cut-off with the same sensitivity (≥23 U/ml), while there was a less marked difference at higher model probability thresholds. At all model thresholds, the PPV was higher than for CA125 cut-offs with equivalent sensitivities.

Table 4. Diagnostic accuracy metrics for a range of Model 1 thresholds and for CA125 at equivalent sensitivities.

	Sens	Spec	PPV	NPV
≥1% model probability	86.4 (81.8-90.2)	89.1 (88.8-89.5)	6.9 (6.1-7.8)	99.9 (99.8-99.9)
CA125 of ≥23 U/ml	86.4 (81.8-90.2)	86.0 (85.6-86.4)	5.5 (4.8-6.2)	99.9 (99.8-99.9)
≥2% model probability	78.5 (73.2-83.2)	94.7 (94.5-95.0)	12.2 (10.8-13.9)	99.8 (99.7-99.8)
Ca125 of ≥35 U/ml	78.5 (73.2-83.2)	94.5 (94.3-94.8)	11.9 (10.4-13.4)	99.8 (99.7-99.8)
≥3% model probability	75.6 (70.2-80.5)	96.9 (96.7-97.1)	18.5 (16.3-20.9)	99.8 (99.7-99.8)
CA125 of ≥39 U/ml	75.6 (70.2-80.5)	95.6 (95.3-95.8)	13.8 (12.1-15.7)	99.8 (99.7-99.8)

The potential implications of applying different model threshold to the study cohort of 29,962 women are illustrated in the schema provided in online supplemental appendix 5. In comparison to the current CA125 cut-off, applying a probability threshold of ≥1%, would result in an additional 1,622 women being identified for further evaluation for ovarian cancer of whom 22 (1.4%) would have ovarian cancer i.e. an additional 1 in every 74 women identified for further evaluation would have ovarian cancer. Applying a ≥3% model probability threshold, instead of the current CA125 cut-off, would result in 706 fewer women being identified for further evaluation of whom 8 (1.1%) would have ovarian cancer. Applying a 2% model threshold, instead of the current CA125 threshold, would result in 58 fewer women being identified for further evaluation of whom none would have ovarian cancer.

DISCUSION

A model consisting of CA125 and age alone demonstrated excellent discrimination and calibration for the identification of ovarian cancer in women presenting to primary care with relevant symptoms. Including additional baseline risk factors, symptom type and routine blood test results did not improve model performance. While the model AUC was only slightly higher than that of CA125 alone, at a fixed sensitivity Model 1 showed superior

specificity and PPV at a range of thresholds. When a ≥1% probability threshold was applied to our cohort, rather than the current CA125 cut-off (≥35 U/ml), one in 74 of the additional women identified by the model had ovarian cancer.

Strengths and limitations

The dataset used in this study was a key strength. Information on candidate variables was identified from a large routinely collected primary care data source, which is broadly representative of the UK general population, and outcome data was obtained from NCRAS (the English cancer registry), which is considered the gold standard source for cancer diagnostic information for epidemiological research.[13] However, use of routinely collected data limited the candidate variables which we could include in our model. For example, family history of ovarian cancer is an established risk factor for the disease but was not included as a candidate variable as it is not routinely recorded in primary care records. Including family history may have introduced bias, as a GP might be more likely to ask about and record family history when there is a strong suspicion of ovarian cancer.

While CA125 testing is only indicated in women with relevant symptoms in UK primary care, symptoms are not always coded within the GP notes; instead, they may be recorded within the free text which is not available for research purposes.[41] In order to include symptoms as predictor variables we had to exclude 19,691 women who are likely to have had relevant symptoms which were simply not coded – this reduced sample size. Although CPRD is one of the largest primary care datasets in the world, our sample size was limited by available data on CA125 testing.

We internally validated models – applying a cross-validation approach – to assess for model optimism and found that this was minimal. However, external validation in an independent dataset is warranted before the models are used in clinical practice.

Comparison with existing literature

In a recent systematic review, we did not identify any other diagnostic prediction models which combined CA125 with symptom variables.[10] Existing primary care diagnostic prediction models, such as QCancer Ovarian,[34] were developed in general primary care

populations (which included both women *with* and *without* symptoms) with the aim of identifying higher risk women for tests, such as CA125, whereas the models in this study were developed within an entirely symptomatic CA125 tested population. Similarly, secondary care models such as the Risk of Ovarian Malignancy Algorithm (ROMA) and the Risk of Malignancy Index (RMI) are not comparable as they were developed in populations all of whom were known to have a pelvic mass, with the aim of distinguishing between benign and malignant masses.[42,43] We have previously reported on the diagnostic accuracy of CA125, as used in the primary care population, and have estimated the probability of ovarian cancer based on patient age and CA125 level,[8] but this is the first study to evaluate the diagnostic accuracy of CA125 predicated models within the symptomatic primary care population.

Baseline patient risk factors, routine blood test results and the type of symptoms with which patients present have previously been found to be predictive of ovarian cancer in general practice,[10] and so were included as candidate variables in this study. However, our results indicate that, in symptomatic women undergoing CA125 testing, the CA125 level is the dominant predictor – the inclusion of other variables, with the exception of age, does not materially improve model performance.

Implications for research and practice

In 2015, NICE recommended urgent cancer referral for symptomatic primary care patients when the risk of a particular cancer reached 3%.[15] However, specific guidance on ovarian cancer investigation and referral was not brought into line with this threshold. NICE recommend that symptomatic women with CA125 levels ≥35 U/ml should be referred by their GP for an ultrasound; only if the ultrasound shows evidence of possible cancer do they qualify for an urgent cancer pathway referral. As primary care ultrasound usually take several weeks to be performed in England,[44] this could delay ovarian cancer diagnosis. The model developed in this study could be used to select women for urgent cancer pathway referral, in line with the NICE 3% threshold, thereby helping to ensure that those at higher risk of undiagnosed cancer receive prompt specialist investigation, diagnosis and treatment.

We recently reported that women with false negative CA125 results in English primary care took twice as long, following testing, to be diagnosed with ovarian cancer as women with abnormal CA125 results.[45] Such delays in diagnosis could have a detrimental effect on patient outcomes, including morbidity and survival. Model 1 would allow women with 'low risk but not no risk' of ovarian cancer to be identified and offered non-urgent evaluation or interval re-assessment. Applying a ≥1% probability threshold (instead of the current CA125 cut-off) this approach could help detect one extra ovarian cancer for every 74 additional patients identified by the model for further evaluation. This evaluation could involve re-testing for CA125, as most women who are retested in primary care prior to diagnosis have rising CA125 levels,[45] or a referral for a non-urgent transvaginal ultrasound. An example of a risk-stratified approach using model thresholds is illustrated in online supplemental appendix 6. As Model 1 is based solely of CA125 level and age, it could readily be incorporated within laboratory computer systems, and the patient's cancer probability reported to the GP alongside the CA125 level.

Implementing a two-tier risk-stratified approach is likely to result in more non-urgent investigation in primary care (*'low risk but not no risk'* women) and more urgent cancer referrals (higher risk women). Any such change in guidelines would require a full health economic evaluation, to assess the potential impact of such a strategy on the healthcare service and on patients, as most of those investigated would ultimately not be diagnosed with ovarian cancer. This evaluation would also be valuable in ensuring that the most appropriate model thresholds are chosen for use in clinical practice. The model thresholds evaluated within this study are of particular relevance to the healthcare system in England. However, following validation in appropriate local data-sets, our model could be used to select women for further investigation for ovarian cancer in line with any regional or national threshold.

In the current study, we have focussed on optimising the initial testing step within the ovarian cancer diagnostic pathway. However, the timely diagnosis of cancer also depends on the accuracy of subsequent testing steps, most notably ultrasound. Further research is needed to ensure that use of model-based risk thresholds improves the accuracy of the diagnostic pathway as a whole.

CONCLUSION

A model consisting of age and CA125 level performs well for the detection of ovarian cancer in symptomatic women in English primary care. A risk-based triage system, informed by this model, has the potential to expedite the diagnosis of ovarian cancer in those at high risk of undiagnosed ovarian cancer (through urgent specialist investigation) and 'low risk but not no risk' of ovarian cancer (through interval retesting or routine ultrasound). Further research is needed to evaluate the practical impact of implementing such an approach on patients and the healthcare system.

Ethical approval

The study was approved by the Independent Scientific Advisory Committee (ISAC) for the Medicines and Healthcare products Regulatory Agency (protocol number 18_184).

Funding

This research arises from the CanTest Collaborative, which is funded by Cancer Research UK [C8640/A23385], of which GF is Clinical Research Fellow, GA is the Senior Statistician, and WH and FMW are Directors. The study was also funded by the National Institute for Health Research (NIHR) School for Primary Care Research (FR17 424). EJC is supported through the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). The funders of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care or Cancer Research UK.

Data sharing

This study is based on data provided by the CPRD and the NCRAS and is subject to a licence agreement that prohibits sharing outside the research team. All data are available through CPRD. Data access is subject to approval from the Independent Scientific Advisory Committee for the Medicines and Healthcare products Regulatory Agency.

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to acknowledge that this work uses data provided by patients and collected by the NHS as part of their care and support.

References

1 Cancer Research UK. Ovarian cancer incidence statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Zero (accessed March 5, 2021).

2 Cancer Research UK. Ovarian cancer survival statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three (accessed April 22, 2021).

3 Jacobs IJ, Menon U, Ryan A, *et al.* Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* 2016; **387**: 945–56.

4 Pinsky PF, Yu K, Kramer BS, *et al.* Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. *Gynecol Oncol* 2016; **143**: 270–5.

5 Barrett J, Sharp DJ, Stapley S, Stabb C, Hamilton W. Pathways to the diagnosis of ovarian cancer in the UK: A cohort study in primary care. *BJOG* 2010; **117**: 610–4.

6 National Cancer Intellegence Network. Routes to Diagnosis 2006-2016 by year, V2.1a. 2019. http://www.ncin.org.uk/publications/routes_to_diagnosis (accessed May 21, 2020).

7 National Collaborating Centre for Cancer. Ovarian Cancer: The recognition and initial management of ovarian cancer (NICE Clinical Guidelines No. 122). Cardiff (UK), 2011.

8 Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. *PLoS Med* 2020; **17**: e1003295.

9 Olaitan A. Recognition and initial management of ovarian cancer: summary of NICE guidance (Rapid response). *BMJ* 2011; **342**: d2073.

10 Funston G, Hardy V, Abel G, *et al.* Identifying Ovarian Cancer in Symptomatic Women: A Systematic Review of Clinical Tools. *Cancers (Basel)* 2020; **12**: 3686.

11 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**: 827–36.

12 Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; **46**: 1093–1093i.

13 Henson KE, Elliss-Brookes L, Coupland VH, *et al.* Data Resource Profile: National Cancer Registration Dataset in England. *Int J Epidemiol* 2019; published online April 23.

14 Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol* 2019; **34**: 91–9.

15 National Institute for Health and Care Excellence. Suspected cancer: recognition and referral (NG12). 2015.

16 Ebell MH, Culp MB, Radke TJ. A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer. *Am J Prev Med* 2016; **50**: 384–94.

17 FIGO Committee on Gynecologic Oncology. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *Int J Gynaecol Obstet* 2014; **124**: 1–5.

18 Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO Classification of Tumours of Female Reproductive Organs. Fourth Edition. Lyon: International Agency for

Research on Cancer, 2014.
19 Watson J, Mounce L, Bailey SER, Cooper SL, Hamilton W. Blood markers for cancer. *BMJ* 2019; **367**: l5774.
20 Bailey SER, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: A prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract* 2017; **67**: e405–13.
21 Watson J, Salisbury C, Banks J, Whiting P, Hamilton W. Predictive value of inflammatory markers for cancer diagnosis in primary care: a prospective cohort study using electronic health records. *Br J Cancer* 2019; **120**: 1045–51.
22 Mathur R, Bhaskaran K, Chaturvedi N, *et al.* Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf)* 2014; **36**: 684–92.
23 National Cancer Registration and Analysis Service. Cancer incidence and survival by major ethnic group. 2009. http://www.ncin.org.uk/search/ethnic+group (accessed April 20, 2021).
24 Cancer Research UK. Ovarian cancer incidence by age. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One (accessed March 5, 2021).
25 Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. *Cancer Biol Med* 2017; **14**: 9–32.
26 World Cancer Research Fund / American Institute for Cancer Research. Diet, nutrition, physical activity and ovarian cancer. 2018. https://www.wcrf.org/sites/default/files/Ovarian-cancer-report.pdf (accessed April 20, 2021).
27 Aune D, Rosenblatt DAN, Chan DSM, *et al.* Anthropometric factors and ovarian cancer risk: A systematic review and nonlinear dose-response meta-analysis of prospective studies. *Int J Cancer* 2015; **136**: 1888–98.
28 Beral V, Hermon C, Peto R, *et al.* Ovarian cancer and body size: Individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS Med* 2012; **9**: e1001200.
29 Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol* 2011; **12**: 785–94.
30 Liu Z, Zhang TT, Zhao JJ, *et al.* The association between overweight, obesity and ovarian cancer: A meta-analysis. *Jpn J Clin Oncol* 2015; **45**: 1107–15.
31 Bhaskaran K, Douglas I, Forbes H, Dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: A population-based cohort study of 5·24 million UK adults. *Lancet* 2014; **384**: 755–65.
32 Stewart LM, Spilsbury K, Jordan S, *et al.* Risk of high-grade serous ovarian cancer associated with pelvic inflammatory disease, parity and breast cancer. *Cancer Epidemiol* 2018; **55**: 110–6.
33 Merriel SWD, Carroll R, Hamilton F, Hamilton W. Association between unexplained hypoalbuminaemia and new cancer diagnoses in UK primary care patients. *Fam Pract* 2016; **33**: 449–52.
34 Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *BMJ* 2012; **344**: d8009–d8009.

35 Peres LC, Mallen AR, Townsend MK, *et al.* High levels of C-reactive protein are associated with an increased risk of ovarian cancer: Results from the ovarian cancer Cohort Consortium. *Cancer Res* 2019; **15**: 5442–51.
36 Harrell FE. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. New York: Springer, 2001.
37 Stata. Impute missing values using chained equations. Stata Manual. https://www.stata.com/manuals13/mimiimputechained.pdf (accessed Dec 14, 2020).
38 Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Med Res Methodol* 2009; **9**: 57.
39 Banks J, Hollinghurst S, Bigwood L, Peters TJ, Walter FM, Hamilton W. Preferences for cancer investigation: A vignette-based study of primary-care attendees. *Lancet Oncol* 2014; **15**: 232–40.
40 Luque-Fernandez MA, Redondo-Sánchez D, Maringe C. cvauroc: Command to compute cross-validated area under the curve for ROC analysis after predictive modeling for binary outcomes. *Stata J* 2019; **19**.
41 Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case-control study. *BMJ Open* 2016; **6**: e011664.
42 Jacobs I, Oram D, Fairbank J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *BJOG* 1990; **100**: 927–31.
43 Moore RG, McMeekin DS, Brown AK, *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009; **112**: 40–6.
44 Nicholson BD, Oke JL, Rose PW, Mant D. Variation in direct access to tests to investigate cancer: A survey of english general practitioners. *PLoS One* 2016; **11**: e0159725.
45 Funston G, Mounce L, Price SJ, *et al.* CA125 test result, test-to-diagnosis interval and stage at diagnosis in ovarian cancer: a cohort study using electronic health records. *Br J Gen Pract* 2021; : BJGP.2020.0859.

Appendix B: ISAC approvals

Contents:

ISAC APPLICATION FORM

PROTOCOLS FOR RESEARCH USING THE CLINICAL PRACTICE RESEARCH DATALINK (CPRD)

For ISAC use only		
Protocol No. Submission date (DD/MM/YYYY)	....	***IMPORTANT*** *Please refer to the **guidance** for '**Completing the ISAC application form**' found on the CPRD website (www.cprd.com/isac). If you have any queries, please contact the ISAC Secretariat at isac@cprd.com.*

SECTION A: GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY

1. Study Title[§] (*Please state the study title below)*

The role of CA125 in the detection of ovarian cancer in symptomatic primary care patients

[§]*Please note: This information will be published on the CPRD's website as part of its transparency policy.*

2. Has any part of this research proposal or a related proposal been previously submitted to ISAC?

Yes[*] ☐ No ☒

**If yes, please provide the previous protocol number/s below. Please also state in your current submission how this/these are related or relevant to this study.*

3. Has this protocol been peer reviewed by another Committee? (e.g. grant award or ethics committee)

Yes[*] ☐ No ☒

**If Yes, please state the name of the reviewing Committee(s) below and provide an outline of the review process and outcome as an Appendix to this protocol :*

4. Type of Study (please tick all the relevant boxes which apply)

Adverse Drug Reaction/Drug Safety	☐	Drug Effectiveness	☐
Drug Utilisation	☐	Pharmacoeconomics	☐
Disease Epidemiology	☒	Post-authorisation Safety	☐
Health care resource utilisation	☒	Methodological Research	☐
Health/Public Health Services Research	☒	Other[*]	☒

[*]*If Other, please specify the type of study here and in the lay summary below:*
Test diagnostic accuracy study

5. Health Outcomes to be Measured[§]

[§]*Please note: This information will be published on CPRD's website as part of its transparency policy.*

Please summarise below the primary/secondary health outcomes to be measured in this research protocol:

Primary outcome:

- Ovarian cancer diagnosis

Secondary outcomes:

- Diagnosis of a cancer other than ovarian
- Histological type and morphology, stage, size, grade of ovarian cancer at diagnosis
- Death from ovarian cancer or death from another cancer

6. Publication: This study is intended for (please tick all the relevant boxes which apply)**:**

Publication in peer-reviewed journals	☒	Presentation at scientific conference	☒
Presentation at company/institutional meetings	☒	Regulatory purposes	☐
Other*	☐		

**If Other, please provide further information:*

SECTION B: INFORMATION ON INVESTIGATORS AND COLLABORATORS

7. Chief Investigator§

Please state the full name, job title, organisation name & e-mail address for correspondence - see guidance notes for eligibility. Please note that there can only be one Chief Investigator per protocol.

Dr Fiona Walter, Principal Researcher in Primary Care Cancer Research, University of Cambridge,
REDACTION: Email removed for confidentiality.

§*Please note: The name and organisation of the Chief Investigator and will be published on CPRD's website as part of its transparency policy*

CV has been previously submitted to ISAC	☐	**CV number:**
A new CV is being submitted with this protocol	☐	
An updated CV is being submitted with this protocol	☒	

8. Affiliation of Chief Investigator (full address)

Primary Care Unit
Department of Public Health and Primary Care
Strangeways Research Laboratory
2 Worts' Causeway
Cambridge CB1 8RN

9. Corresponding Applicant§

Please state the full name, affiliation(s) and e-mail address below:

Dr Garth Funston, Department of Public Health and Primary Care, REDACTION: Email removed for confidentiality.

§*Please note: The name and organisation of the corresponding applicant and their organisation name will be published on CPRD's website as part of its transparency policy*

Same as chief investigator	☐		
CV has been previously submitted to ISAC	☒	**CV number:**	234_18
A new CV is being submitted with this protocol	☐		
An updated CV is being submitted with this protocol	☐		

10. List of all investigators/collaborators§

Please list the full name, affiliation(s) and e-mail address* of all collaborators, other than the Chief Investigator below:

§*Please note: The name of all investigators and their organisations/institutions will be published on CPRD's website as part of its transparency policy*

Other investigator: Professor Willie Hamilton,
Exeter Medical School, University of Exeter, REDACTION: Email removed for confidentiality.

CV has been previously submitted to ISAC	☒	**CV number:**	201_15CEPS
A new CV is being submitted with this protocol	☐		
An updated CV is being submitted with this protocol	☐		

Other investigator: Dr Emma Crosbie
Department of Gynaecological Oncology, University of Manchester, REDACTION: Email removed for confidentiality.

CV has been previously submitted to ISAC	☐	**CV number:**
A new CV is being submitted with this protocol	☒	
An updated CV is being submitted with this protocol	☐	

Other investigator: Dr Gary Abel
Exeter Medical School, University of Exeter, REDACTION: Email removed for confidentiality.
CV has been previously submitted to ISAC ☒ **CV number:** 149_18
A new CV is being submitted with this protocol ☐
An updated CV is being submitted with this protocol ☐

Other investigator:
CV has been previously submitted to ISAC ☐ **CV number:**
A new CV is being submitted with this protocol ☐
An updated CV is being submitted with this protocol ☐

[Please add more investigators as necessary]

Please note that your ISAC application form and protocol **must be copied to all e-mail addresses listed above at the time of submission of your application to the ISAC mailbox. Failure to do so will result in delays in the processing of your application.*

11. Conflict of interest statement*
Please provide a draft of the conflict (or competing) of interest (COI) statement that you intend to include in any publication which might result from this work

We have no conflicts of interest to declare.

**Please refer to the International Committee of Medical Journal Editors (ICMJE) for guidance on what constitutes a COI.*

12. Experience/expertise available
Please complete the following questions to indicate the experience/ expertise available within the team of investigators/collaborators actively involved in the proposed research, including the analysis of data and interpretation of results.

	Previous GPRD/CPRD Studies	**Publications using GPRD/CPRD data**
None	☐	☐
1-3	☐	☐
> 3	☒	☒

Experience/Expertise available	**Yes**	**No**
Is statistical expertise available within the research team? *If yes, please indicate the name(s) of the relevant investigator(s)* Dr Gary Abel	☒	☐
Is experience of handling large data sets (>1 million records) available within the research team? *If yes, please indicate the name(s) of the relevant investigator(s)* Professor Willie Hamilton, Dr Gary Abel	☒	☐
Is experience of practising in UK primary care available to or within the research team? *If yes, please indicate the name(s) of the relevant investigator(s)* Dr Garth Funston, Dr Fiona Walter, Professor Willie Hamilton	☒	☐

13. References relating to your study
Please list up to 3 references (most relevant) relating to your proposed study:

1) National Institute of Clinical Excellence. Suspected cancer: recognition and referral. www.nice.org.uk/guidance/ng12. Accessed 26 Jul 2017.
2) Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of an algorithm. BMJ. 2011;344:d8009–d8009.
3) Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ. doi: 10.1136/bmj.i3139

SECTION C: ACCESS TO THE DATA

14. Financial Sponsor of study[§]

[§]*Please note: The name of the source of funding will be published on CPRD's website as part of its transparency policy*

Pharmaceutical Industry	☐	*Please specify name and country:*
Academia	☐	*Please specify name and country:*
Government / NHS	☐	*Please specify name and country:*
Charity	☒	*Please specify name and country:* Cancer Research UK
Other	☐	*Please specify name and country:*
None	☐	

15. Type of Institution conducting the research

Pharmaceutical Industry	☐	*Please specify name and country:*
Academia	☒	*Please specify name and country:* University of Cambridge, UK
Government Department	☐	*Please specify name and country:*
Research Service Provider	☐	*Please specify name and country:*
NHS	☐	*Please specify name and country:*
Other	☐	*Please specify name and country:*

16. Data access arrangements

The financial sponsor/ collaborator* has a licence for CPRD GOLD and will extract the data	☐
The institution carrying out the analysis has a licence for CPRD GOLD and will extract the data**	☒
A data set will be provided by the CPRD[¥€]	☒
CPRD has been commissioned to extract the data and perform the analyses[€]	☐
Other:	☐

If Other, please specify:

**Collaborators supplying data for this study must be named on the protocol as co-applicants.*
***If data sources other than CPRD GOLD are required, these will be supplied by CPRD*
[¥]*Please note that datasets provided by CPRD are limited in size; applicants should contact CPRD (enquiries@cprd.com) if a dataset of >300,000 patients is required.*
[€]*Investigators must discuss their request with a member of the CPRD Research team before submitting an ISAC application. Please contact the CPRD Research Team on +44 (20) 3080 6383 or email (enquiries@cprd.com) to discuss your requirements. Please also state the name of CPRD Research team with whom you have discussed this request (provide the date of discussion and any relevant reference information):*

Name of CPRD Researcher Helen Booth Reference number (where available) Date of contact 1/3/18

17. Primary care data

Please specify which primary care data set(s) are required)

Vision only (Default for CPRD studies ☒ Both Vision and EMIS®* ☐
EMIS® only* ☐

Note: Vision and EMIS are different practice management systems. CPRD has traditionally collected data from Vision practice. Data collected from EMIS is currently under evaluation prior to wider release.
Investigators requiring the use of EMIS data **must discuss the study with a member of the CPRD Research team before submitting an ISAC application*

Please state the name of the CPRD Researcher with whom you have discussed your request for EMIS data:
Name of CPRD Researcher Reference number (where available) Date of contact

18. Site Location of Data

a) Processing location(s):

Location area - UK / EEA / Worldwide:

UK

Organisation address:

Primary Care Unit
Department of Public Health and Primary Care
Strangeways Research Laboratory
2 Worts' Causeway
Cambridge CB1 8RN
Note: Please enter the location details of where the data for this study will be used (processed).

b) Storage Location(s)

Location area - UK / EEA / Worldwide:

UK

Organisation address:
Primary Care Unit
Department of Public Health and Primary Care
Strangeways Research Laboratory
2 Worts' Causeway
Cambridge CB1 8RN
Note: Please enter the location details of where the data for this study will be stored.

c) Territory of analysis - UK / EEA / Worldwide:
UK

Primary Care Unit
Department of Public Health and Primary Care
Strangeways Research Laboratory
2 Worts' Causeway
Cambridge CB1 8RN

Note: Please enter the details of where the data for this study will be analysed.

SECTION D: INFORMATION ON DATA LINKAGES

19. Does this protocol seek access to linked data

Yes* ☒ No☐ If No, please move to section E.

Research groups which have not previously accessed CPRD linked data resources **must discuss access to these resources with a member of the CPRD Research team, before submitting an ISAC application. Investigators requiring access to HES Accident and Emergency data, HES Diagnostic Imaging Dataset, PROMS data, the Pregnancy Register, Cancer Registration, SACT and CPES data and the Mental Health Services Data Set **must** also discuss this with a member of the CPRD Research team before submitting an ISAC application. Please contact the CPRD Research Team on +44 (20) 3080 6383 or email enquiries@cprd.comto discuss your requirements **before** submitting your application.*

Please state the name of the CPRD Researcher with whom you have discussed your linkage request.

Name of CPRD Researcher Helen Booth Reference number (where available) Date of contact 1/3/18

Please note that as part of the ISAC review of linkages, your protocol may be shared - in confidence - with a representative of the requested linked data set(s) and summary details may be shared - in confidence - with the Confidentiality Advisory Group of the Health Research Authority.

20. Please select the source(s) of linked data being requested§
§Please note: This information will be published on the CPRD's website as part of its transparency policy.

☒ ONS Death Registration Data
☒ HES Admitted Patient Care
☐ HES Outpatient
☐ HES Accident and Emergency
☒NCRAS (National Cancer Registration and Analysis Service) Cancer Registration Data *
☐ NCRAS Cancer Patient Experience Survey (CPES) data*
☐ NCRAS Systemic Anti-Cancer Treatment (SACT) data*

☐ HES Diagnostic Imaging Dataset ☐ Mental Health Services Data Set (MHDS)
☐ HES PROMS (Patient Reported Outcomes Measure)**
☐ CPRD Mother Baby Link
☐ Pregnancy Register

☐ Practice Level Index of Multiple Deprivation (Standard)
☐ Practice Level Index of Multiple Deprivation (Bespoke)
☐ Patient Level Index of Multiple Deprivation***
☒ Patient Level Townsend Score ***

**Applicants seeking access to NCRAS data must complete a Cancer Dataset Agreement form (available from CPRD). This should be submitted to the ISAC as an appendix to your protocol. Please also note that applicants seeking access to cancer registry data must provide consent for publication of their study title and study institution on the UK Cancer Registry website.*
***Assessment of the quality of care delivered to NHS patients in England undergoing four procedures: hip replacement, knee replacement, groin hernia and varicose veins. Please note that patient level PROMS data are only available for non-commercial purposes, such as academic research, or in connection with delivering services to the NHS.*
**** 'Patient level IMD and Townsend scores will not be supplied for the same study*
*****If "Other" is specified, please provide the name of the individual in the CPRD Research team with whom this linkage has been discussed.*

Name of CPRD Researcher Reference number (where available) Date of contact

21. Total number of linked datasets requested <u>including</u> CPRD GOLD

Number of linked datasets requested (*practice/ 'patient' level Index of Multiple Deprivation, Townsend Score, the CPRD Mother Baby Link and the Pregnancy Register should **<u>not</u>** be included in this count)* 4

Please note: Where ≥5 linked datasets are requested, approval may be required from the Confidentiality Advisory Group (CAG) to access these data

22. Is linkage to a <u>local¥</u> dataset with <1 million patients being requested?

Yes* ☐ No ☒

**If yes, please provide further details:*
¥ Data from defined geographical areas i.e. non-national datasets.

23. If you have requested one or more linked data sets, please indicate whether the Chief Investigator or any of the collaborators listed in question 5 above, have access to these data in a patient identifiable form (e.g. full date of birth, NHS number, patient post code), or associated with an identifiable patient index.
Yes* ☐ No ☒

** If yes, please provide further details:*

24. Does this study involve linking to patient *identifiable* data (e.g. hold date of birth, NHS number, patient post code) from other sources?
Yes ☐ No ☒

SECTION E: VALIDATION/VERIFICATION

25. Does this protocol describe a purely observational study using CPRD data?

Yes* ☒ No** ☐

** Yes: If you will be using data obtained from the CPRD Group, this study does not require separate ethics approval from an NHS Research Ethics Committee.*
*** No: You may need to seek separate ethics approval from an NHS Research Ethics Committee for this study. The ISAC will provide advice on whether this may be needed.*

26. Does this protocol involve requesting any additional information from GPs?

Yes* ☐ No ☒

** If yes, please indicate what will be required:*

Completion of questionnaires by the GP[ψ] Yes ☐ No ☐
Is the questionnaire a validated instrument? Yes ☐ No ☐
If yes, has permission been obtained to use the instrument? Yes ☐ No ☐
Please provide further information:

Other (please describe)

[ψ] Any questionnaire for completion by GPs or other health care professional must be approved by ISAC before circulation for completion.

27. Does this study require contact with patients in order for them to complete a questionnaire?

Yes* ☐ No ☒

**Please note that any questionnaire for completion by patients must be approved by ISAC before circulation for completion.*

28. Does this study require contact with patients in order to collect a sample?

Yes* ☐ No ☒

** Please state what will be collected:*

SECTION F: DECLARATION

29. Signature from the Chief Investigator

- I have read the guidance on '***Completion of the ISAC application form***' and '***Contents of CPRD ISAC Research Protocols***' and have understood these;
- I have read the submitted version of this research protocol, including all supporting documents, and confirm that these are accurate.
- I am suitably qualified and experienced to perform and/or supervise the research study proposed.
- I agree to conduct or supervise the study described in accordance with the relevant, current protocol
- I agree to abide by all ethical, legal and scientific guidelines that relate to access and use of CPRD data for research
- I understand that the details provided in sections marked with (§) in the application form and protocol will be published on the CPRD website in line with CPRD's transparency policy.
- I agree to inform the CPRD of the final outcome of the research study: publication, prolonged delay, completion or termination of the study.

Name: Dr Fiona M Walter Date: 29th June 2018 e-Signature (type name): Fiona M Walter

PROTOCOL INFORMATION REQUIRED

The following sections below **must** be included in the CPRD ISAC research protocol. Please refer to the guidance on '***Contents of CPRD ISAC Research Protocols***' (www.cprd.com/isac) for more information on how to complete the sections below. Pages should be numbered. All abbreviations must be defined on first use.

Applicants must complete all sections listed below **Sections which do not apply should be completed as '*Not Applicable*'**
A. Study Title§ §*Please note: This information will be published on CPRD's website as part of its transparency policy* The role of CA125 in the detection of ovarian cancer in symptomatic primary care patients
B. Lay Summary (Max. 200 words)§ §*Please note: This information will be published on CPRD's website as part of its transparency policy* Ovarian cancer is the 5th most common cause of cancer related death in UK women. The majority of women are diagnosed late and only 46 out of every 100 UK women survive for 5 years after diagnosis. Early diagnosis is likely to result in better patient outcomes including survival. However, early diagnosis is challenging. The symptoms of ovarian cancer are vague and the same symptoms occur in non-worrying medical conditions, so it is can be difficult for GPs to decide which patients need to be sent to hospital urgently for more tests and which can be reassured. Simple blood tests, such as CA125, can be used to help GPs make these decisions. However, we don't know how good CA125 is when used in primary care or what 'cut-off point' to use for an abnormal result. In this study, we aim to determine how effective CA125 is at picking up cancer in women visiting their GP with symptoms which could be caused to ovarian cancer, and identify the most appropriate abnormal CA125 cut-off. This work will help GPs to make decisions regarding investigation and referral of symptomatic women.
C. Technical Summary (Max. 200 words)§ §*Please note: This information will be published on CPRD's website as part of its transparency policy* Ovarian cancer has the worst prognosis of any gynaecological cancer. Early diagnosis is likely to improve survival, and, while symptoms occur in all stages, they are also common in benign conditions. Tests are needed to help distinguish malignant from benign disease in symptomatic patients. The serum biomarker CA125 is frequently elevated in women with ovarian cancer. It is used as a first line investigation in primary care, in the UK and internationally, in patients presenting with symptoms that might be caused by ovarian cancer. Despite widespread use, the diagnostic accuracy of CA125 in the primary care population has not been established and the current 'abnormal threshold' (35u/ml) is not based on primary care data. In this prospective cohort study, we will determine the diagnostic accuracy (sensitivity, specificity, PPV, NPV) of CA125 in a symptomatic primary care population and identify CA125 thresholds that equate to a range of risk thresholds (PPVs). As CA125 levels and ovarian cancer risk are influenced by patient variables, we will produce stratified thresholds based on key variables e.g. age. This work will allow GPs to make decisions about further investigation and referral based on patient risk.

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

D. Objectives, Specific Aims and Rationale

Study objective: To evaluate the diagnostic accuracy of CA125 in a symptomatic primary care population and develop primary care evidenced thresholds to guide further investigation and referral.

Aim 1: To determine the **overall** diagnostic accuracy of CA125 at the current threshold and identify a range of CA125 thresholds equating to different PPV's
Rationale: CA125 is advocated as the first line test in patients presenting to primary care in the UK (and several other countries) with symptoms suggestive of ovarian cancer. The National Institute of Clinical Excellence (NICE) has set a 'risk threshold' of 3% for investigation and cancer pathway referrals in symptomatic patients [1]. Research indicates that patients would opt for investigation at much lower risk thresholds [2]. We will determine the diagnostic accuracy (sensitivity, specificity, PPV, NPV) of CA125 in UK primary care at its current threshold and calculate thresholds that equate to a range of PPVs (including 1% and 3%). This will aid GPs to make informed decisions about further investigations and referrals. NICE guidance may be revised in 2019/2020, and our results would greatly facilitate guideline revision.

Aim 2: To determine the diagnostic accuracy of CA125 at the current threshold and identify a range of primary care evidenced thresholds for **patient subgroups**
Rationale: CA125 levels are affected by a number of patient variables such as age and co-morbidities. Further, some of the same factors affect the risk of ovarian cancer. Thus, the same CA125 level may equate to different PPVs in different patient subgroups e.g. <50 yrs old >50yrs old. As such, different CA125 thresholds may be required in different groups to reach the same ovarian cancer 'risk threshold'. We will determine the diagnostic performance of CA125 in specific subgroups and calculate thresholds for these subgroups which equate to a range of PPVs (including 1% and 3%). This will aid GPs to make more individualised decisions about further investigations and referrals.

Aim 3: To examine how CA125 testing, using the current threshold, impacts on the stage of diagnosis for women with values close to the current threshold
Rationale: Current guidance advocates that women with CA125 levels above 35u/ml undergo further investigation. As the test is not perfect there will be women below the threshold who have, as yet undiagnosed, ovarian cancer. It is likely that these women will go on to be diagnosed, but in the intervening period the stage of disease may have advanced. It is hoped that by diagnosing women with an elevated CA125 level earlier than might have been done without the test, that they are diagnosed at an earlier stage which may allow curative treatment and better survival. Women just above and just below the threshold have no tangible difference in their risk of ovarian cancer. However, the timeliness of diagnosis is likely to be impacted by which side of the line they fall. We will exploit this arbitrary boundary, to establish the effect of initiating investigations and subsequent follow up on the basis of a positive test result using a regression discontinuity design.

E. Study Background

Ovarian cancer has the worst prognosis of any form of gynaecological cancer and accounts for over 4000 deaths in the UK each year. Survival is stage dependant and the majority of women are not diagnosed until the disease is advanced, which contributes to the UK's poor 5-year survival rate of 46% [3, 4]. Most patients with ovarian cancer who present in primary care have common non-specific symptoms e.g. bloating. Testing in primary care can help determine which patients are at greatest risk of cancer and warrant referral, and which can be reassured.

CA125 is a high molecular weight glycoprotein of unknown function expressed by several normal human tissues. Serum CA125 levels are raised in ovarian cancer and several other malignant and benign conditions [5–7]. Levels can also vary during the menstrual cycle and in pregnancy [7]. In 2011, NICE published guidelines advocating CA125 testing in primary care patients presenting with symptoms that might represent ovarian cancer [8]. Further investigation was recommended in patients with a CA125 level above 35u/ml. Following this guidance, there was a substantial increase in CA125 testing [9]. Several other countries also advocate CA125 as an initial test for ovarian cancer in primary care [10, 11].

Despite its widespread use, the diagnostic performance of CA125 in patients presenting to primary care with symptoms suggestive of ovarian cancer has not been determined. The 'abnormal threshold' of 35u/ml, which is currently employed in both primary and secondary care throughout the world, is derived from a 1983 study in which

1% of 888 healthy patient and 82% (n=101) of patients with ovarian cancer had a CA125 level above 35u/ml [5]. CA125 has subsequently been studied extensively in secondary care, however, this research cannot readily be translated to the primary care setting as the characteristics of the population, including the incidence of ovarian cancer and other benign and malignant conditions which affect CA25 levels, is inherently different. These differences result in spectrum bias or spectrum effect, where the performance of a test varies depending on the population in which it is used [12].

In addition, use of a single CA125 threshold may not be the optimal approach in primary care, as cancer risk and CA125 levels vary between patient groups. For example, CA125 levels are significantly higher in groups of apparently healthy women under the age of 50 years than groups 50 years and over [13], while the incidence of ovarian cancer is lower in women under the age of 50 years [14]. As such, a single threshold may equate to different PPVs for ovarian cancer in different patient groups. Thresholds stratified on the basis of key variables, such as age, can be more accurate in determining disease status [15].

NICE recommends that in symptomatic patients presenting to primary care, a 3% risk (PPV) of ovarian cancer is sufficient to trigger urgent investigation or a cancer pathway referral. Patients have indicated that they would opt for investigation at lower risk thresholds e.g. 1%. In this study, we will determine what CA125 thresholds equate to different risk thresholds. The work will help guide GPs when making decisions about further investigations and referrals.

Definitions and Terminology

The term ovarian cancer can be used to encompass a number of distinct diseases which, while occurring in a similar anatomical region, differ in their tissue of origin, aetiology, molecular pathogenesis, clinical behaviour, presentation, treatment and prognosis. Ovarian cancer can be broadly divided into epithelial (>90%) and non-epithelial in origin. Non-epithelial cancers are a heterogeneous group of rare cancers that present early, have a relatively good prognosis and generally do not cause elevation in serum CA125, while epithelial cancers present late, have a poor prognosis and usually result in elevated CA125.
The term ovarian cancer is a misnomer. Epithelial ovarian cancer can arise from the epithelial lining of the ovary, fallopian tube (most common site) or the peritoneum [16]. Ovarian, fallopian tube and primary peritoneal cancers are now classified and staged using the same systems [17], and are treated collectively in current NICE guidance covering recognition and initial management.
In this protocol, the term 'ovarian cancer' will be used to describe **ovarian, fallopian tube and primary peritoneal cancer**.

F. Study Type

This is a diagnostic accuracy study in which we seek to determine the diagnostic performance of CA125 in a symptomatic primary care population. We believe this falls into the 'hypothesis testing' category as outlined in the ISAC protocol guidance.

G. Study Design

We will use a prospective cohort design to determine the diagnostic accuracy of CA125 when used in UK general practice.

H. Feasibility counts

Patients with a CA125 result:
We used the Define tool to estimate the numbers of patients with a new CA125 code and the number of patients with both a new CA125 code and a new code for ovarian cancer during our period of interest. We used Read codes for CA125 and epithelial ovarian cancer (Appendix). We applied the following restrictions-
-Date- 1st May 2011 to 1st June 2016
-Age at index date- ≥18 years at index date
-Gender- Female
-System- vision
-CA125 first ever code in study period
-Ovarian cancer first ever code in study period

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

This search identified **121,012 patients** with a new code for CA125 and **1423 patients** with a new code for CA125 and a new code for ovarian cancer.

Given the differences between epithelial and non-epithelial cancer (discussed in Definitions and Terminology) we will perform a sub analysis using diagnosis of epithelial ovarian cancer as the outcome. As codes for ovarian cancer within CPRD are non-specific, it is not possible to determine which CA125 tested patients have epithelial ovarian cancer on the basis of CPRD codes. As cancer registry data includes histological cancer type we will perform this subgroup analysis on patients with linked cancer registry data, which is around 57% of patients [18]. As 90% of ovarian cancer is epithelial in origin we anticipate that this analysis will include at least **68,977** CA125 tested patients of whom an estimated **730** patients will have a cancer registry documented epithelial ovarian cancer.

I. Sample size considerations

The principal objective of this study is to determine the diagnostic accuracy of CA125 when used within a symptomatic primary care population. As no study has evaluated the sensitivity and specificity of CA125 in primary care, we have used values from a large meta-analysis performed in secondary care in patients with a known pelvic mass (sensitivity 79%, specificity 78%) to perform the below calculations [19]. All calculations were performed using Stata 15.1.

With a sample consisting of 121,012 CA125 tested patients and 1423 patients with ovarian cancer, we estimate 95% confidence intervals of 76.8% to 81.1% around a sensitivity of 79%, 77.8% to 78.2% around a specificity of 78% and 3.9% to 4.3% around a PPV of 4.1%.

These confidence intervals are narrow and will allow for precision in our calculations of CA125 diagnostic accuracy for the overall population.

The second objective of this study is to determine the diagnostic accuracy of CA125 in key patient subgroups. e.g. different age groups, different ethnicities and patients with specific ovarian cancer associated symptoms. The numbers of CA125 tested women and the number of patients with ovarian cancer within each of these groups is unknown but is likely to be small for some groups. For example, we are interested in evaluating the diagnostic performance of CA125 in patients of different ethnicities as baseline levels are believed to vary between different ethnic groups [20]. In the 2011 Census of England and Wales, 3.4% of the population were classified as Black [21]. Extrapolating this to our sample, we anticipate that 4,114 of the CA125 tested women will be of Black ethnicity. This will provide 80% power to see a difference in ovarian cancer diagnoses rates between 1.7% and 2.3% in non-Black/Black women. We would also have 90% power to see a change in sensitivity from 79% to 92% in this group. Changes in either the prevalence of ovarian cancer in the tested population, the sensitivity or specificity of the test will lead to changes in the PPV. Restricting estimates of diagnostic accuracy to Black women will result in broader confidence intervals, which would be expected to be as follows if the test characteristics did not change: sensitivity (65.0-89.5%), specificity (76.7 -79.3%) and PPV (2.9-5.5%). In order to maximise our ability to evaluate test diagnostic accuracy in these subgroups we have requested data for all CA125 tested patients during our period of interest.

J. Data Linkage Required (if applicable):[§]

[§]*Please note that the data linkage/s requested in research protocols will be published by the CPRD as part of its transparency policy*

Cancer Registry- Diagnosis of ovarian cancer is the primary outcome in the study. Although concordance between CPRD and the cancer registry is high, additional cases can be identified from the registry data [22] and CPRD codes are frequently non-specific. The cancer registry contains information on cancer stage, grade and tumour size at diagnosis and histological cancer type, which will be included in the study as discussed below.

HES Admitted Patient Care (integrated data)- ethnicity is a variable in the study. Ethnicity is more frequently recorded within HES than CPRD data [23].

Office of National Statistics (ONS) Deaths Registration data- Death due to ovarian cancer will be included as a secondary outcome. This linkage is required to cross validate the cause and date of death.

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

K. Study population

The study population will consist of women who underwent a CA125 test between the 1st of May 2011 and a date 2 years before the study commences (chosen to allow 2 years of follow-up for all patients). Read codes for CA125 (listed in the appendix) will be used to define our cohort.

Inclusion criteria:
-Women ≥18 years of age at the time of the first CA125 test during the study period.
-No recorded ovarian cancer diagnosis (either within the cancer registry or CPRD data) at the time of the first CA125 test.
- ≥ 1 year of up to standard CPRD records prior to first CA125 testing.

Exclusion criteria:
-Women < 18 years of age at the time of the first CA125 test during the study period.
-Women with a recorded ovarian cancer diagnosis (either within the cancer registry or CPRD data) at the time of CA125 testing during the study period. We will not exclude patients with other comorbidities known to affect CA125 levels as we wish our study cohort to remain representative of the population in which the test is being performed.

L. Selection of comparison group(s) or controls

Internal comparison- we will compare the incidence of ovarian cancer in patients with / without an elevated CA125 level.

M. Exposures, Health Outcomes§ and Covariates
§Please note: Summary information on health outcomes (as included on the ISAC application form above)will be published on CPRD's website as part of its transparency policy

Primary outcome variable-
A new diagnosis of ovarian cancer recorded within the clinical record / referral files or linked cancer registry data, within 2 years of a CA125 test.
A code list for ovarian cancer (appendix), will be used to search the dataset. A follow-up period of two years from the date of initial CA125 testing has been chosen as a compromise between picking up all related cancers (which would be maximised by using a longer follow-up) and picking up unrelated cancers (which would be minimised by using a shorter follow-up). We will also stratify cancer diagnoses by the number of days the diagnosis occurred following CA125 testing.

Secondary outcome variables-
a) A new diagnosis of a cancer other than ovarian cancer, recorded within the clinical record / referral files or linked cancer registry data, within 2 years of a CA125 test. CA125 may be elevated in a number of other cancers e.g. endometrial and lung. Validated code lists for cancers developed by Professor Hamilton's group will be used.
b) Death, from 1) ovarian cancer, 2) any cancer, as recorded in CPRD data or ONS death registration data within 2 years of a CA125 test.
c) Stage of cancer as recorded in cancer registry data. It is acknowledged that this information will only be available for patients with linked cancer registry data.
d) Tumour morphology and histology, as recorded in the cancer registry. CA125 levels are known to vary by histological type [7].
e) Tumour size, as recorded in cancer registry.
f) Tumour grade as recorded in cancer registry.

Principle explanatory variable-
CA125 level.
CA125 tests will be identified from CPRD data using Read codes (appendix).

Other variables-
CA125 level and / or ovarian cancer risk are affected by a number of variables such as symptoms, comorbidities, lifestyle factors and family / patient history of cancer. We will seek to examine a number of variables including:

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

Variable	Source	Comment / rationale
Symptoms and signs		
Symptoms and signs within 1 month of CA125 testing	Validated code lists for symptoms related to ovarian cancer in CPRD	Risk of ovarian cancer is greater in patients presenting to primary care with certain symptoms [24]. In addition, symptoms may be related to an underlying condition other than cancer e.g. endometriosis, which may affect CA125 level.
Patient characteristics		
Age at time of CA125 testing	CPRD data	CA125 levels are thought to be higher in groups of younger women than older women [13]. Ovarian cancer risk is age related.
Ethnicity	CPRD data and, where available, HES data. Codes for the 16 ethnic groups recognised in the 2001 census	For analysis, these codes will be collapsed into 4 ethnic groups- 'white', 'mixed', 'Asian or Asian British', 'Chinese or other'. CA125 level is thought to vary between groups of women of different ethnicities [20].
Parity	Relevant Read codes within CPRD	CaA125 levels are thought to be lower in groups of parous vs nonporous women [25].
Tests		
Test results including FBC (platelet count, total white cell count, haemoglobin, platelet count), GFR, creatinine, CRP, albumin	Read codes for test results within CPRD	Some results e.g. GFR and CRP may indicate underlying conditions which affect CA125 levels. Other tests e.g. platelet count, may be predictive of ovarian cancer [26].
Comorbidities and operations		
Comorbidities (fibroids, endometriosis, ovarian cyst, renal failure, pre-existing cancer other than ovarian) recorded within 2 years of CA125 testing	Relevant Read codes within CPRD	Comorbidities known to affect CA125 level will be included. The risk of ovarian cancer is greater in patients with endometriosis [27].
Personal and family history of cancer and cancer syndromes		
Personal history of breast, endometrial, stomach, colon, small intestine, hepatobiliary, urinary tract, brain or skin cancer	Relevant Read codes within CPRD	Risk of ovarian cancer is likely to be greater in women who have had a BRCA or Lynch syndrome related cancer
Family history of breast, ovarian, endometrial, stomach, colon, small intestine, hepatobiliary, urinary tract, brain or skin cancer in first degree relatives	Relevant Read codes within CPRD	Risk of ovarian cancer may be greater in women with a family member who has had a BRCA or Lynch syndrome related cancer
BRCA mutations or Lynch syndrome	Relevant Read codes within CPRD	Risk of ovarian cancer is greater in women with a BRCA mutation or Lynch syndrome
BRCA mutations or Lynch syndrome in a first degree family member	Relevant Read codes within CPRD	Risk of ovarian cancer may be greater in women with a family member who has a BRCA mutation or Lynch syndrome

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

N. Data/ Statistical Analysis

Firstly, we will use descriptive statistics to summarise outcomes (e.g. numbers and proportions of patients diagnosed with ovarian cancer and other cancers) and variables (e.g. numbers and proportions of patients with symptom codes). All variables will be predefined prior to data analysis. An exploratory phase will assess the association between these pre-defined variables and CA125 level and predefined variables and primary outcome using regression analysis.

The principal objective of this study is to determine the diagnostic accuracy of CA125 in symptomatic primary care patients. In order to do this we will calculate the number of true positive, false positive, true negative and false negative CA125 results. From this, we will calculate the diagnostic accuracy (sensitivity, specificity, PPV and NPV) of CA125. 95% confidence intervals will be calculated for the various measures of test accuracy. This analysis will be repeated for secondary outcomes.

An important objective of the study is to determine what CA125 thresholds equate to different risk thresholds for ovarian cancer. As such, we will construct ROC curves to illustrate CA125 performance at a range of thresholds equating to various PPVs.

We expect that the diagnostic accuracy of CA125 may differ between different patient groups e.g. >50/<50 years old. We will use logistic regression to establish which factors influence the risk of having cancer, and explore whether the association between having cancer and CA125 level varies between predefined patient groups. If there is evidence of either of these associations it will suggest different test characteristics. Once the important factors have been identified, we will repeat the ROC curve analysis for these key patient subgroups.

As discussed above, CPRD codes for ovarian cancer are non-specific and will not allow us to distinguish epithelial ovarian cancer from non-epithelial types of ovarian cancer. A subgroup analysis will be performed using data for patients with linked CPRD-cancer registry data where histological cancer type can be accurately determined. We are including malignant and borderline tumours as our primary endpoint while excluding in situ lesions. Sensitivity analyses will be performed, using the morphology variable in the cancer registry, firstly excluding borderline tumours then including in situ and borderline tumours.

Our study aims to determine the performance of CA125 in symptomatic patients. However, real world use of CA125 in UK general practice may differ from intended use in symptomatic patients as set out in NICE guidelines. To test this, we will perform a sensitivity analysis using data from patients with CPRD codes for symptoms included in NICE guidance.

Finally we will perform a regression discontinuity design analysis. Regression discontinuity designs exploit arbitrary thresholds at which a certain action is taken. This has recently been used to demonstrate that the use of PSA testing in a screening population leads to an overdiagnosis of early stage prostate cancers with no change in overall mortality [28]. In this study we will use a regression discontinuity design to examine changes in early stage (TNM stages 1 and 2) and in late stage (TNM stage 3 and 4) cancers as well as overall diagnosis rates at the currently used CA125 threshold. We do not expect to see evidence of overdiagnosis and thus no change in overall diagnosis rates, combined with increases in early stage diagnoses above the threshold would support the use of CA125 testing to facilitate earlier diagnosis. Following the previous work on PSA testing we will use the user written regression discontinuity Stata module [29].

Analyses will be performed using STATA version 15.1.

O. Plan for addressing confounding

This study will assess the diagnostic accuracy of CA125 as it is used in real world UK general practice. CPRD data is largely representative of the UK general practice population.

Nevertheless, we have endeavoured to identify variables that are associated with CA125 level and ovarian cancer risk, as outlined above. We will explore the relationship between each variable, CA125 level and cancer incidence using regression analysis. Wherever possible, we will produce variable specific ROC curves.

P. Plans for addressing missing data

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

We will describe the extent of missing data and use appropriate approaches to evaluate its impact.

Previous studies have found that the recording of blood results within CPRD data is excellent. Linkage to cancer registry data provides us with two sources from which we can identify cancer diagnoses. As such, we anticipate having relatively complete data for our principal explanatory variable and outcomes. As with any cohort or diagnostic accuracy study, there is a risk of loss to follow-up e.g. if a patient moves away. However, we do not anticipate that the loss to follow-up rate will be significantly different between CA125 positive/negative patients prior to diagnosis.

For several of our variables and secondary outcomes we have identified more than one source of information e.g. mortality (CPRD, ONS). This will help reduce the impact of missing data.

While we anticipate having relatively complete data for our principal explanatory variable and primary outcome, we recognise that data for several other variables, e.g. personal and family history of cancer may be less complete. Any such data will be interpreted with caution and the limitations of the data will be highlighted in any related publications. Some secondary outcomes e.g. stage at diagnosis, are also likely to have some missing data. However, missing outcome data will not result in bias in estimated associations under the Missing At Random (MAR) assumption.

Q. Patient or user group involvement (if applicable)

Our group has a patient and public involvement representative, Mrs Margaret Johnson. She has given valuable input into this proposal and the protocol and will continue to be involved throughout the study.

R. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

We aim to publish this work in peer reviewed journals. In addition, it will be presented at national and international primary care and cancer conferences. This work will form an integral part of Garth Funston's PhD thesis.
Previous routine data studies have led to the development and dissemination of paper and online risk tools [30]. We will explore the possibility of converting our findings into an e-algorithm / tool which can be used by GPs in the assessment of patients.

S. Limitations of the study design, data sources, and analytic methods

Several minor limitations exist.

One limitation is that we will be unable to conclusively identify the reasons for CA125 requests. However, we will be able to identify symptoms and signs recorded in CRPD data close to the time of CA125 testing, which may suggest a rationale for requests. As we cannot say for certain why CA125 tests were ordered, we will be unable to make firm judgements about the appropriateness of testing or whether it was conducted in line with NICE guidelines. Despite this, we will be able to demonstrate how CA125 performs as is it is currently used in UK general practice and a sensitivity analysis will be performed using data from patients with CPRD codes for symptoms included in NICE guidance.

Reliance on Read codes to identify symptoms and the presence of other variables is a limitation, as coding may be incomplete.

We have selected a 2 year follow up from the point of CA125 testing. While we believe that this is an appropriate period, it is possible that incidental ovarian cancers may occur and be diagnosed during this time or that patients with false negative CA125 results may not have re-presented and been diagnosed.

We recognise that the number of cancer patients within some of the subgroups may be limited, which in turn will limit the precision of our estimates of diagnostic accuracy for these groups. In order to maximise our ability to evaluate test diagnostic accuracy in these subgroups we have requested data for all CA125 tested patients during our period of interest.

T. References

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

1. National Institute of Clinical Excellence. Suspected cancer: recognition and referral. www.nice.org.uk/guidance/ng12. Accessed 26 Jul 2017.
2. Banks J, Hollinghurst S, Bigwood L, et al. Preferences for cancer investigation: A vignette-based study of primary-care attendees. Lancet Oncol. 2014;15:232–240.
3. Cancer Research UK. Ovarian cancer mortality statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/mortality. Accessed 26 Jul 2017.
4. Maringe C, Walters S, Butler J, et al. Stage at diagnosis and ovarian cancer survival: Evidence from the International Cancer Benchmarking Partnership. Gynecol Oncol. 2012;127:75–82.
5. Bast RC, Klug TL, John E St., et al. A Radioimmunoassay Using a Monoclonal Antibody to Monitor the Course of Epithelial Ovarian Cancer. N Engl J Med. 1983;309:883–887.
6. Molina R, Augé JM, Bosch X, et al. Usefulness of serum tumor markers, including progastrin-releasing peptide, in patients with lung cancer: correlation with histology. Tumour Biol. 2009;30:121–9.
7. Jacobs I, Bast RC. The CA 125 tumour-associated antigen: a review of the literature. Hum Reprod. 1989;4:1–12.
8. National Institute of Clinical Excellence. Ovarian cancer: recognition and initial management. https://www.nice.org.uk/guidance/cg122/resources/ovarian-cancer-recognition-and-initial-management-35109446543557. Accessed 14 Feb 2018.
9. Giamougiannis P, Ebef NG dit, Mahmood T. The implications of new guideline 122 On the detection of ovarian cancer. Eur J Obstet Gynecol Reprod Biol. 2016;206:e154.
10. Ireland National Cancer Control Programme. Ovarian cancer GP referral for symptomatic women. www.hse.ie/eng/services/list/5/cancer/profinfo/resources/gpreferrals/Ovarian-Cancer-Referral-Guidelines.pdf. Accessed 2 Dec 2017.
11. Cancer Australia. Assessment of symptoms that may be ovarian cancer: a guide for GPs. In: 2015. https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/assessment-symptoms-may-be-ovarian-cancer-guide-gps. Accessed 14 Feb 2018.
12. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ. doi: 10.1136/bmj.i3139
13. Haga Y, Sakamoto K, Egami H, Yoshimura R, Akagi M. Evaluation of serum CA125 values in healthy individuals and pregnant women. Am J Med Sci. 1986;292:25–9.
14. Cancer Research UK. Ovarian cancer incidence statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One. Accessed 14 Feb 2018.
15. Quinn MM, Kao CN, Ahmad AK, et al. Age-stratified thresholds of anti-Müllerian hormone improve prediction of polycystic ovary syndrome over a population-based threshold. Clin Endocrinol (Oxf). 2017;87:733–740.
16. Karnezis AN, Cho KR, Gilks CB, Pearce CL, Huntsman DG. The disparate origins of ovarian cancers: Pathogenesis and prevention strategies. Nat Rev Cancer. 2017;17:65–74.
17. Meinhold-Heerlein I, Fotopoulou C, Harter P, et al. The new WHO classification of ovarian, fallopian tube, and primary peritoneal cancer and its clinical implications. Arch Gynecol Obstet. 2016;293:695–700.

Applicants must complete all sections listed below
Sections which do not apply should be completed as '*Not Applicable*'

18. Clinical Practice Research Datalink. CPRD linked data. https://www.cprd.com/dataAccess/linkeddata.asp. Accessed 25 Apr 2018.
19. Ferraro S, Braga F, Lanzoni M, et al. Serum human epididymis protein 4 vs. carbohydrate antigen 125 for ovarian cancer diagnosis: A systematic review. Biochim Clin. 2013;37:179–189.
20. Johnson CC, Kessel B, Riley TL, et al. The epidemiology of CA-125 in women without evidence of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial. Gynecol Oncol. 2008;110:383–389.
21. Gov.UK. Ethnicity facts and figures. https://www.ethnicity-facts-figures.service.gov.uk/ethnicity-in-the-uk/population-by-ethnicity. Accessed 25 Apr 2018.
22. Boggon R, Van Staa TP, Chapman M, et al. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. Pharmacoepidemiol Drug Saf. 2013;22:168–175.
23. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J Public Health (Oxf). 2014;36:684–692.
24. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. BMJ. 2009;339:b2998.
25. Lowe KA, Shah C, Wallace E, et al. Effects of personal characteristics on serum CA125, mesothelin, and HE4 levels in healthy postmenopausal women at high-risk for ovarian cancer. Cancer Epidemiol Biomarkers Prev. 2008;17:2480–2487.
26. Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. BMJ. 2011;344:d8009–d8009.
27. Kvaskoff M, Mu F, Terry KL, et al. Endometriosis: A high-risk population for major chronic diseases? Hum Reprod Update. 2014;21:500–516.
28. Jonathan Shoag, Joshua Halpern, Brian Eisner et al. Efficacy of Prostate-Specific Antigen Screening: Use of Regression Discontinuity in the PLCO Cancer Screening Trial. JAMMA Oncol. 2015;1:984–986.
29. Nichols A. RD: Stata module for regression discontinuity estimation. https://ideas.repec.org/c/boc/bocode/s456888.html. Accessed 27 Jun 2018.
30. Hamilton W, Green T, Martins T, et al. Evaluation of risk assessment tools for suspected cancer in general practice: A cohort study. Br J Gen Pract. doi: 10.3399/bjgp13X660751

List of Appendices *(Submit all appendices as separate documents to this application)*

Appendix 1: CA125 codes
Appendix 2: Ovarian cancer codes

Applicants must complete all sections listed below Sections which do not apply should be completed as '*Not Applicable*'

Dear ISAC Secretariat,

RE: Protocol 18_184: "The role of CA125 in the detection of ovarian cancer in symptomatic primary care patients"

We would like to request the following minor amendments to the above approved protocol:

1) Amendment 1: Rather than accepting a code for ovarian cancer in CPRD or in NCRAS we will only accept an ovarian cancer record in NCRAS data. This will necessitate only using patients with NCRAS linkage for this aspect of the study. There are 2 reasons for this change:
   -Concordance between CPRD and NCRAS data is variable with regards to cancer recording [1]. A paper, published after our ISAC submission, highlights that solely relying on CPRD for cancer case identification could affect the results of diagnostic accuracy studies, such as ours, due to the potential for misclassification bias [1]. While compiling a list of Read codes for this study we found a number of nonspecific codes which may or may not indicate the presence of ovarian cancer- this could lead to misclassification bias. NCRAS reports a near 100% case ascertainment, collects data from multiple sources and is generally considered the gold standard for cancer recording in population cancer research and so we feel it is appropriate to use NCRAS to identify cancer cases in our study.
   -As noted in the protocol, we wish to perform a sensitivity analysis excluding borderline ovarian tumours- this is not possible using CPRD cancer codes due to their non-specific nature (morphology and topographical NCRAS codes are needed).

The sample should still be adequate to determine diagnostic accuracy of CA125 with a high degree of precision: 72,182 patients have linkage in the dataset and, based on our feasibility count, we would assume that up to 849 women to have ovarian cancer. Using estimated sensitivity (79%) and specificity (78%) from the sample size considerations section of our protocol, we would anticipate narrow confidence intervals around measures of diagnostic accuracy (particularly our main measure which is PPV): PPV 4.1% (95% CI: 3.8-4.4), NPV 99.7% (95% CI:99.6-99.7), sensitivity 79% (95% CI: 76.1-81.7), specificity 78% (95% CI: 77.7-78.3). These are very similar to the CI estimates in the approved ISAC protocol.

2) Amendment 2: We will use a 1 year rather than a 2 year follow-up period post CA125 testing. While some studies use a 2 year period the majority of similar recent studies (several of which have been published since we submitted our original ISAC protocol) have used a 1 year period [2-3]. We believe that a 1 year period will be sufficient for the majority of ovarian cancers to be diagnosed following CA125 testing and will minimise the number of incidental ovarian cancers included in our analysis.

3) Amendment 3: In the protocol we state that we will exclude patients without 1 year of up-to-standard follow-up prior to CA125 testing. We wish to change this to exclude patients whose practices are not up-to-standard at the point of CA125 testing (the focus of our study is activity post CA125 testing rather than prior to CA125 testing).

4) Amendment 4: We wish to add the following exclusion criteria- "Patients with a CA125 record prior to the study start date but within 12 months of the initial CA125 test during the study period". This is because evidence has emerged since we submitted the ISAC protocol that the PPV of a repeat test may be different to that of an initial test [2].

These changes do not fundamentally alter the design or aim of the study and appear to fall under the minor amendment category in your document "Guidance on Resubmissions and Amendments of ISAC Research Protocols". A protocol (with these amendments) will be submitted with the paper when we come to publish our results.

Please do not hesitate to contact me if you require further information.

Best wishes,

REDACTED.
Signature redacted
for confidentiality.

Fiona Walter

Study Chief Investigator

References

[1] Badrick E, Renehan I, Renehan AG. Linkage of the UK Clinical Practice Research Datalink with the national cancer registry. Eur J Epidemiol. 2019;34:101-102.
[2] Watson J, Salisbury C, Banks J, Whiting P, Hamilton W. Predictive value of inflammatory markers for cancer diagnosis in primary care: a prospective cohort study using electronic health records. BJC. 2019;120:1045–1051.
[3] Watson J, Jones, H, Banks J, Whiting P, Salisbury C, Hamilton W. Use of multiple inflammatory marker tests in primary care: using Clinical Practice Research Datalink to evaluate accuracy**.** 2019;69:e462-e469.

Appendix C: Code lists

This appendix comprises code lists used in this thesis.

Contents:

Supplementary Table C.1. Read codes and terms used to identify CA125 tested women.

Medcode	Read code	Read term
14565	44a6.00	CA125 level
108230	44a6000	Serum CA 125 (cancer antigen 125) level
9228	44a1.00	Carbohydrate antigen 125 level

Supplementary Table C.2. Categorisation of ICD-O morphology and behaviour codes in the baseline dataset.

ICD-O morphology and behaviours	Morphology	Thesis category	Invasive	Borderline
8000/1	Neoplasm, uncertain whether benign or malignant	Unknown	0	1
8000/3	Neoplasm, malignant	Unknown	1	0
8010/1	Epithelial tumour, uncertain behaviour	Epithelial (other)	0	1
8010/3	Carcinoma NOS	Epithelial (unknown)	1	0
8020/3	Carcinoma, undifferentiated NOS	Epithelial (other)	1	0
8041/3	Small cell carcinoma NOS	Epithelial (other)	1	0
8070/3	Squamous cell carcinoma NOS	Epithelial (other)	1	0
8140/3	Adenocarcinoma NOS	Epithelial (unknown)	1	0
8240/3	Carcinoid tumor NOS	Non-epithelial	1	0
8246/3	Neuroendocrine carcinoma	Epithelial (other)	1	0
8310/3	Clear cell adenocarcinoma NOS	Clear cell	1	0
8313/1	Clear cell adenofibroma of borderline malignancy	Clear cell	0	1
8323/3	Mixed cell adenocarcinoma	Epithelial (other)	1	0
8380/3	Endometrioid carcinoma	Endometrioid	1	0
8440/3	Cystadenocarcinoma NOS	Epithelial (unknown)	1	0
8441/1	Serous cystadenoma, borderline malignancy	Serous	0	1
8441/3	Serous cystadenocarcinoma NOS	Serous	1	0
8442/3	Serous cystadenoma, borderline malignancy	Serous	0	1

8442/5	Serous cystadenoma, microinvasion	Serous	0	1
8460/3	Papillary serous cystadenocarcinoma	Serous	1	0
8461/3	Serous surface papillary carcinoma	Serous	1	0
8462/3	Papillary serous cystadenoma, borderline malignancy	Serous	0	1
8462/5	Papillary serous cystadenoa, microinvasion	Serous	0	1
8470/3	Mucinous cystadenocarcinoma NOS	Mucinous	1	0
8472/3	Mucinous cystadenoma, borderline malignancy	Mucinous	0	1
8472/5	Mucinous cystadenoma, microinvasive	Mucinous	0	1
8473/3	Papillary mucinous cystadenoma, borderline malignancy	Mucinous	0	1
8480/3	Mucinous adenocarcinoma	Mucinous	1	0
8620/1	Granulosa cell tumor NOS	Non-epithelial	1	0
8620/3	Granulosa cell tumor, malignant	Non-epithelial	1	0
8631/3	Sertoli-Leydig cell tumour, poorly differentiated	Non-epithelial	1	0
8890/3	Leiomyosarcoma NOS	Non-epithelial	1	0
8950/3	Mullerian mixed tumor	Epithelial (other)	1	0
8980/3	Carcinosarcoma, NOS	Epithelial (other)	1	0
9014/1	Serous adenofibroma of borderline malignancy	Serous	0	1
9014/3	Serous adenocarcinofibroma	Serous	1	0
9015/1	Mucinous adenofibroma of borderline malignancy	Mucinous	0	1
9071/3	Endodermal sinus tumour	Non-epithelial	1	0
9080/3	Teratoma, malignant NOS	Non-epithelial	1	0
9085/3	Mixed germ cell tumor	Non-epithelial	1	0

0 = no, 1= Yes

Supplementary Table C.3. Ovarian cancer symptom Read code list.

Read code	Read description	Mapped NICE symptom
25C..11	O/E - epigastric pain on palp.	Abdo/pelvic pain
R090A00	[D]Pain in left iliac fossa	Abdo/pelvic pain
R090600	[D]Umbilical pain	Abdo/pelvic pain
25CZ.00	O/E -abd.pain on palpation NOS	Abdo/pelvic pain
R090y00	[D]Other specified abdominal pain	Abdo/pelvic pain
197..14	Subcostal pain	Abdo/pelvic pain
R090M00	[D]Right lower quadrant pain	Abdo/pelvic pain
197A.00	Generalised abdominal pain	Abdo/pelvic pain
197Z.00	Site of GIT pain NOS	Abdo/pelvic pain
1973.00	Left subcostal pain	Abdo/pelvic pain
R090N00	[D]Nonspecific abdominal pain	Abdo/pelvic pain
25E..00	O/E - rebound tenderness	Abdo/pelvic pain
1963.00	Non-colicky abdominal pain	Abdo/pelvic pain
25D..11	O/E - guarding of abdomen	Abdo/pelvic pain
25DA.00	O/E - guarding - L.iliac	Abdo/pelvic pain
R090C00	[D]Loin pain	Abdo/pelvic pain
25C7.00	O/E - abd. pain - L.lumbar	Abdo/pelvic pain
N33A000	Bony pelvic pain	Abdo/pelvic pain
R090100	[D]Abdominal colic	Abdo/pelvic pain
R090K00	[D]Left upper quadrant pain	Abdo/pelvic pain
25C..15	O/E - abdomen tender	Abdo/pelvic pain
2I18100	Tenderness of epigastrium	Abdo/pelvic pain
Ryu1100	[X]Other and unspecified abdominal pain	Abdo/pelvic pain
197..12	Iliac fossa pain	Abdo/pelvic pain
25C8.00	O/E - abd. pain - R.iliac	Abdo/pelvic pain
196..12	Type of GIT pain - symptom	Abdo/pelvic pain
R090E00	[D]Recurrent acute abdominal pain	Abdo/pelvic pain
197..00	Site of GIT pain	Abdo/pelvic pain
196..11	Abdominal pain type	Abdo/pelvic pain
25D2.00	O/E - guarding-R.hypochondrium	Abdo/pelvic pain
R090.00	[D]Abdominal pain	Abdo/pelvic pain
25EZ.00	O/E - rebound tenderness NOS	Abdo/pelvic pain
25C..14	O/E - umbilical pain on palp.	Abdo/pelvic pain
1979.00	Suprapubic pain	Abdo/pelvic pain
1975.00	Left flank pain	Abdo/pelvic pain
25D3.00	O/E - guarding - epigastrium	Abdo/pelvic pain
Ryu1000	[X]Pain localized to other parts of lower abdomen	Abdo/pelvic pain
25C3.00	O/E - abd. pain - epigastrium	Abdo/pelvic pain
25C..12	O/E - iliac pain on palpation	Abdo/pelvic pain
R090G00	[D]Pelvic and perineal pain	Abdo/pelvic pain
25C5.00	O/E - abd. pain - R.lumbar	Abdo/pelvic pain
1A59.00	C/O pelvic pain	Abdo/pelvic pain
197B.00	Upper abdominal pain	Abdo/pelvic pain

25D..00	O/E - guarding on palpation	Abdo/pelvic pain
25D9.00	O/E - guarding - hypogastrium	Abdo/pelvic pain
25C6.00	O/E - abd. pain - umbilical	Abdo/pelvic pain
1978.00	Left iliac fossa pain	Abdo/pelvic pain
2I18.00	O/E - tenderness/pain	Abdo/pelvic pain
1829.00	Retrosternal pain	Abdo/pelvic pain
R090000	[D]Abdominal tenderness	Abdo/pelvic pain
25D6.00	O/E - guarding - umbilical	Abdo/pelvic pain
R090L00	[D]Left lower quadrant pain	Abdo/pelvic pain
R090H00	[D]Upper abdominal pain	Abdo/pelvic pain
25C2.00	O/E - abd.pain-R.hypochondrium	Abdo/pelvic pain
2I18.12	O/E - tenderness	Abdo/pelvic pain
R090P00	[D]Functional abdominal pain syndrome	Abdo/pelvic pain
25C..13	O/E - lumbar pain on palpation	Abdo/pelvic pain
1969.00	Abdominal pain	Abdo/pelvic pain
1962.00	Colicky abdominal pain	Abdo/pelvic pain
1972.00	Epigastric pain	Abdo/pelvic pain
R090z00	[D]Abdominal pain NOS	Abdo/pelvic pain
197..13	Site of abdominal pain	Abdo/pelvic pain
1977.00	Right iliac fossa pain	Abdo/pelvic pain
197..11	Flank pain	Abdo/pelvic pain
25C9.00	O/E - abd. pain - hypogastrium	Abdo/pelvic pain
1969000	Abdominal wall pain	Abdo/pelvic pain
R090200	[D]Colic NOS	Abdo/pelvic pain
25CA.00	O/E - abd. pain - L.iliac	Abdo/pelvic pain
R090F00	[D]Acute abdomen	Abdo/pelvic pain
25DZ.00	O/E -guarding on palpation NOS	Abdo/pelvic pain
25D8.00	O/E - guarding - R.iliac	Abdo/pelvic pain
K58y000	Other pelvic pain - female	Abdo/pelvic pain
1976.00	Right flank pain	Abdo/pelvic pain
197D.00	Right upper quadrant pain	Abdo/pelvic pain
E278000	Psychogenic pain unspecified	Abdo/pelvic pain
1DC5.00	Griping pain	Abdo/pelvic pain
25C4.00	O/E - abd.pain-L.hypochondrium	Abdo/pelvic pain
R090B00	[D]Groin pain	Abdo/pelvic pain
R090J00	[D]Right upper quadrant pain	Abdo/pelvic pain
196..00	Type of GIT pain	Abdo/pelvic pain
R090G12	[D] Perineal pain	Abdo/pelvic pain
R090800	[D]Suprapubic pain	Abdo/pelvic pain
197A.11	General abdominal pain-symptom	Abdo/pelvic pain
25D4.00	O/E - guarding-L.hypochondrium	Abdo/pelvic pain
196Z.00	Type of GIT pain NOS	Abdo/pelvic pain
R090500	[D]Epigastric pain	Abdo/pelvic pain
197C.00	Lower abdominal pain	Abdo/pelvic pain
R090400	[D]Abdominal cramps	Abdo/pelvic pain
R090700	[D]Hypochondrial pain	Abdo/pelvic pain

25F..00	O/E - abdominal rigidity	Abdo/pelvic pain
1968.00	Abdominal discomfort	Abdo/pelvic pain
1971.00	Central abdominal pain	Abdo/pelvic pain
R090900	[D]Pain in right iliac fossa	Abdo/pelvic pain
R096.00	[D]Acute abdomen	Abdo/pelvic pain
R073200	[D]Gas pain (abdominal)	Abdo/pelvic pain
R090G11	[D] Pelvic pain	Abdo/pelvic pain
25C..00	O/E - abdo. pain on palpation	Abdo/pelvic pain
25K2.00	O/E - abdominal mass - hard	Abdo/pelvic mass
2642.00	O/E - pelvic mass palpable-RIF	Abdo/pelvic mass
25J7.00	Right iliac fossa mass	Abdo/pelvic mass
7H2C500	Biopsy of abdominal mass	Abdo/pelvic mass
25J3.00	O/E -abd.mass fills 1 quadrant	Abdo/pelvic mass
R093100	[D]Abdominal mass	Abdo/pelvic mass
25L2.00	O/E -abd.mass -irregular shape	Abdo/pelvic mass
R093700	[D]Umbilical mass	Abdo/pelvic mass
25K..00	O/E-abdominal mass consistency	Abdo/pelvic mass
R093600	[D]Umbilical swelling	Abdo/pelvic mass
2641.00	O/E - pelvic mass palpable-LIF	Abdo/pelvic mass
2688.00	O/E - VE - pelvic mass NOS	Abdo/pelvic mass
R093A00	[D]Groin mass	Abdo/pelvic mass
25NZ.00	O/E -abd.mass -border def. NOS	Abdo/pelvic mass
25J9.00	Epigastric mass	Abdo/pelvic mass
R093800	[D]Umbilical lump	Abdo/pelvic mass
25N..00	O/E - abd.mass -border defined	Abdo/pelvic mass
25M1.00	O/E - abd.mass moves with resp	Abdo/pelvic mass
R093400	[D]Pelvic mass	Abdo/pelvic mass
25KZ.00	O/E - abd.mass consistency NOS	Abdo/pelvic mass
2643.00	O/E - central pelvic mass	Abdo/pelvic mass
25MZ.00	O/E - abd.mass + respn. NOS	Abdo/pelvic mass
R093.00	[D]Swelling, mass or lump within abdomen or pelvis	Abdo/pelvic mass
7H2C600	Biopsy of pelvic mass	Abdo/pelvic mass
25K1.00	O/E - abdominal mass - soft	Abdo/pelvic mass
25M..00	O/E - abd.mass movt.with resp.	Abdo/pelvic mass
R093200	[D]Abdominal lump	Abdo/pelvic mass
25N1.00	O/E -abd.mass-upper border def	Abdo/pelvic mass
25L..00	O/E - abdominal mass shape	Abdo/pelvic mass
R093500	[D]Pelvic lump	Abdo/pelvic mass
25N2.00	O/E -abd.mass-lower border def	Abdo/pelvic mass
264Z.00	O/E - pelvic mass palpable NOS	Abdo/pelvic mass
25J8.00	O/E left lower abdominal mass	Abdo/pelvic mass
25J..00	O/E - abdominal mass palpated	Abdo/pelvic mass
25J5.00	O/E - abd. mass fills abdomen	Abdo/pelvic mass
25K3.00	O/E - abdominal mass-very hard	Abdo/pelvic mass

R093z00	[D]Swelling, mass or lump within abdomen or pelvis NOS	Abdo/pelvic mass
R093000	[D]Abdominal swelling	Abdo/pelvic mass
25M2.00	O/E - abd.mass still with resp	Abdo/pelvic mass
R093111	[D]Lump stomach	Abdo/pelvic mass
264..00	O/E - pelvic mass palpated	Abdo/pelvic mass
R093B00	[D]Groin lump	Abdo/pelvic mass
R030000	[D]Appetite loss	Appetite loss
R030.00	[D]Anorexia	Appetite loss
ZC11.00	Restricting food intake	Appetite loss
E275600	Non-organic loss of appetite	Appetite loss
1612.12	Loss of appetite - symptom	Appetite loss
1615.00	Reduced appetite	Appetite loss
1612.00	Appetite loss - anorexia	Appetite loss
R030z00	[D]Anorexia NOS	Appetite loss
1612.11	Anorexia symptom	Appetite loss
Eu50y12	[X]Psychogenic loss of appetite	Appetite loss
25O3.00	O/E-ascites-fluid thrill shown	Ascites
4F32.00	Ascitic fluid: malignant cells	Ascites
7H2B100	Peritoneal to venous drainage for ascites	Ascites
J56y100	Chronic peritoneal effusion	Ascites
4JL5.00	Peritoneal fluid for organism	Ascites
4F12.00	Ascitic fluid exam. normal	Ascites
4JL4.00	Ascitic fluid for organism	Ascites
R095.00	[D]Ascites	Ascites
4F46.00	Ascitic fluid protein level	Ascites
4F5..00	Ascitic fluid cell count	Ascites
25O2.00	O/E - ascites - dipping shown	Ascites
B576200	Malignant ascites	Ascites
25O..00	O/E - ascites	Ascites
4F2..00	Ascitic fluid appearance	Ascites
7H2B000	Paracentesis abdominis for ascites	Ascites
4F1Z.00	Ascitic fluid gen. exam. NOS	Ascites
4F11.00	Ascitic fluid sent for exam.	Ascites
25O4.00	O/E -ascites-shifting dullness	Ascites
4F2Z.00	Ascitic fluid appearance NOS	Ascites
4F...00	Ascitic fluid examination	Ascites
G86y100	Chylous ascites	Ascites
R095z00	[D]Ascites NOS	Ascites
4F1..00	Ascitic fluid exam. - general	Ascites
4F3..00	Ascitic fluid microscopy	Ascites
25OZ.00	O/E - ascites NOS	Ascites
4JL4.11	Ascitic fluid for C/S	Ascites
4F45.00	Ascitic fluid glucose level	Ascites
R095000	[D]Fluid in peritoneal cavity	Ascites
R125.00	[D]Peritoneal fluid abnormal	Ascites

4F4..00	Ascitic fluid chemistry	Ascites
7H2B200	Drainage of ascites NEC	Ascites
4FZ..00	Ascitic fluid exam. NOS	Ascites
4F43.00	Ascitic fluid lactate dehydrogenase level	Ascites
4F21.00	Ascitic fluid clear	Ascites
4F35.00	Ascitic fluid: organisms	Ascites
7H2B113	Insertion of peritoneal to venous shunt for ascites	Ascites
4F13.00	Ascitic fluid exam. abnormal	Ascites
J4zz.11	Diarrhoea - presumed non-infectious	CIBH
R076z00	[D]Incontinence of faeces NOS	CIBH
25Q6.00	O/E - PR-rectum full of faeces	CIBH
4745.00	Faeces consistency: hard	CIBH
19F..12	Loose stools	CIBH
J4...13	Noninfective diarrhoea	CIBH
19EA.11	Altered bowel habit	CIBH
J503100	Faecal impaction	CIBH
4746.00	Faeces consistency: dry	CIBH
J525.00	Functional diarrhoea	CIBH
19F3.00	Spurious (overflow) diarrhoea	CIBH
19F..00	Diarrhoea symptoms	CIBH
J521000	Irritable bowel syndrome with diarrhoea	CIBH
J520300	Drug induced constipation	CIBH
R078.00	[D]Change in bowel habit	CIBH
J43z.11	Chronic diarrhoea	CIBH
19FZ.00	Diarrhoea symptom NOS	CIBH
6643.00	GIT symptom changes	CIBH
R076.00	[D]Incontinence of faeces	CIBH
8138.00	Removal of impacted faeces	CIBH
J432.11	Allergic diarrhoea	CIBH
19E3.11	Incontinent of faeces symptom	CIBH
J433.11	Dietetic diarrhoea	CIBH
2AF2.00	O/E - defaec.ref.abn.-constip.	CIBH
4752.00	Faeces quantity: bulky	CIBH
19FZ.11	Diarrhoea & vomiting, symptom	CIBH
E264300	Psychogenic diarrhoea	CIBH
J52y100	Difficulty in ability to defaecate	CIBH
19C2.00	Constipated	CIBH
E264500	Psychogenic constipation	CIBH
19E3.00	Incontinent of faeces	CIBH
19G..00	Diarrhoea and vomiting	CIBH
J4z..11	Presumed noninfectious diarrhoea	CIBH
19C..11	Constipation symptom	CIBH
R079.00	[D] Defaecation painful	CIBH
19F2.00	Diarrhoea	CIBH
J520y00	Other specified constipation	CIBH

19F..11	Diarrhoea	CIBH
J520z00	Constipation NOS	CIBH
J520000	Acute constipation	CIBH
7728100	Manual removal of impacted faeces from rectum	CIBH
J520400	Chronic constipation	CIBH
Eu45317	[X]Psychogenic diarrhoea	CIBH
J520200	Chronic constipation without overflow	CIBH
J520.00	Constipation - functional	CIBH
4743.00	Faeces consistency: semi-fluid	CIBH
19C..00	Constipation	CIBH
4744.00	Faeces consistency: fluid	CIBH
19EA.00	Change in bowel habit	CIBH
J520100	Chronic constipation with overflow	CIBH
R077100	[D] Stools loose	CIBH
19CZ.00	Constipation NOS	CIBH
19EC.00	Painful defaecation	CIBH
19A..00	Abdominal distension symptom	Distension
19A2.00	Abdomen feels bloated	Distension
19AZ.00	Abd. distension symptom NOS	Distension
19A3.00	Abdomen feels distended	Distension
19B..12	Bloating symptom	Distension
R073300	[D]Abdominal distension, gaseous	Distension
R073400	[D]Bloating	Distension
Eu46000	[X]Neurasthenia	Fatigue
1684.00	Malaise/lethargy	Fatigue
F286.14	Post-viral fatigue syndrome	Fatigue
2254.00	O/E - apathetic	Fatigue
F286.12	Postviral fatigue syndrome	Fatigue
E205.11	Nervous exhaustion	Fatigue
168Z.00	Tiredness symptom NOS	Fatigue
R007z00	[D]Malaise and fatigue NOS	Fatigue
F286.13	PVFS - Postviral fatigue syn	Fatigue
R007411	[D]Post viral debility	Fatigue
F286.00	Chronic fatigue syndrome	Fatigue
1682.00	Fatigue	Fatigue
1B32.00	Weakness present	Fatigue
168..00	Tiredness symptom	Fatigue
F286000	Mild chronic fatigue syndrome	Fatigue
R007000	[D]Malaise	Fatigue
E205.00	Neurasthenia - nervous debility	Fatigue
168..11	Fatigue - symptom	Fatigue
R007z11	[D]Lassitude	Fatigue
R204.00	[D]Senile exhaustion	Fatigue
F286100	Moderate chronic fatigue syndrome	Fatigue
E205.12	Tired all the time	Fatigue

168..12	Lethargy - symptom	Fatigue
R007300	[D]Lethargy	Fatigue
F286200	Severe chronic fatigue syndrome	Fatigue
1684.11	C/O - debility - malaise	Fatigue
R2y3.00	[D]Debility, unspecified	Fatigue
R007211	[D]General weakness	Fatigue
1683.00	Tired all the time	Fatigue
1683.11	C/O - 'tired all the time'	Fatigue
1688.00	Exhaustion	Fatigue
F286.11	CFS - Chronic fatigue syndrome	Fatigue
R007200	[D]Asthenia NOS	Fatigue
R007500	[D]Tiredness	Fatigue
2832.12	O/E - weakness	Fatigue
R007400	[D]Postviral (asthenic) syndrome	Fatigue
1B3..12	Weakness symptoms	Fatigue
Eu46011	[X]Fatigue syndrome	Fatigue
R007.00	[D]Malaise and fatigue	Fatigue
R007100	[D]Fatigue	Fatigue
R084z00	[D]Frequency of micturition or polyuria NOS	Urinary frequency
1A1..00	Micturition frequency	Urinary frequency
1A1..11	Frequency of micturition	Urinary frequency
R084.00	[D]Micturition frequency and polyuria	Urinary frequency
R084000	[D]Frequency of micturition, unspecified	Urinary frequency
1A1..13	Urinary frequency	Urinary frequency
1A12.00	Frequency of micturition	Urinary frequency
1A1Z.00	Micturition frequency NOS	Urinary frequency
1A25.11	Urgency of micturition	Urinary urgency
1A25.00	Urgency	Urinary urgency
R086200	[D] Urgency of micturition	Urinary urgency
R032.00	[D]Abnormal loss of weight	Weight loss
1623.00	Weight decreasing	Weight loss
22A6.00	O/E - Underweight	Weight loss
1D1A.00	Complaining of weight loss	Weight loss
R034800	[D]Underweight	Weight loss
2287.00	Abnormally thin	Weight loss
1625.11	Abnormal weight loss - symptom	Weight loss
R2y4000	[D]Wasting disease	Weight loss
R2y4.00	[D]Cachexia	Weight loss
1625.00	Abnormal weight loss	Weight loss
R2y4z00	[D]Cachexia NOS	Weight loss
1627.00	Unintentional weight loss	Weight loss
2224.00	O/E - cachexic	Weight loss
E264311	Spurious diarrhoea	CIBH
3930.00	Bowels: incontinent	CIBH
19EE.00	Increased frequency of defaecation	CIBH

19EF.00	Urgent desire for stool	CIBH
2AF3.00	O/E -defaec.ref-spurious diar.	CIBH
25J4.00	O/E - abd. mass fills half abd	Abdo/pelvic mass
25LZ.00	O/E - abd. mass shape NOS	Abdo/pelvic mass
25L1.00	O/E - abd. mass -regular shape	Abdo/pelvic mass
25JZ.00	O/E - abd. mass palpated NOS	Abdo/pelvic mass
25J2.00	O/E - abd. mass < 1 quadrant	Abdo/pelvic mass
25R3.00	O/E - dullness over abd. mass	Abdo/pelvic mass
25R2.00	O/E - tympany over abd. mass	Abdo/pelvic mass
J521.00	Irritable colon - Irritable bowel syndrome	IBS
J521.13	Spastic colon	IBS
J521.11	Irritable bowel syndrome	IBS
J521200	IBS characterised by alternating bowel habit	IBS

I identified and used these additional symptom terms while performing the research described in Chapter 6.

Supplementary Table C.4. BRCA mutation Read code list.

Read code	Read term
J521200	BRCA2 gene mutation positive
4L44.00	BCRA1 gene mutation positive

Supplementary Table C.5. Family history of breast or ovarian cancer Read code list.

Read code	Read term
1243.11	FH: Breast cancer
ZV16300	[V]Family history of malignant neoplasm of breast
124C.00	FH: neoplasm of ovary
1245.11	FH: Ovarian carcinoma

Supplementary Table C.6. Bilateral oophorectomy codes.

Read code	Read term
7E04500	Abdominal hysterectomy and bilateral salpingoophorectomy
7E04H00	Subtotl abdominal hysterectomy & bilat salpingo-oophorectomy
7E04511	Abdominal hysterectomy & bilateral salpingoophorectomy (BSO)
7E04512	TAH - total abdom hysterectomy & bilateral salpingoophorect
7E10000	Bilateral salpingoophorectomy
159B.00	H/O: bilateral oophorectomy
7E10200	Bilateral oophorectomy NEC
7E04P00	Radical hysterectomy with bilateral salpingo-oophorectomy
7E05600	Lap assist vag hysterectomy with bilat salpingo-oophorectomy
7E04900	TAH - Tot abdom hysterectomy and BSO - bilat salpingophorect
7E11100	Salpingoophorectomy remaining solitary fallop tube and ovary
7E04B00	Lapar total abdominal hysterect bilat salpingo-oophorectomy
7E11500	Oophorectomy of remaining solitary ovary NEC

Supplementary Table C.7. Read codes and terms for platelet count.

Read code	Read term
42P4.00	Platelet count abnormal
42P1.00	Platelet count normal
42PZ.00	Platelet count NOS
42P..00	Platelet count

Supplementary Table C.8. Read codes and terms for haemoglobin level.

Read code	Read term
423..00	Haemoglobin estimation
423B.00	Haemoglobin abnormal
4238	Haemoglobin borderline high
423A.00	Haemoglobin very high
4234	Haemoglobin very low
4236	Haemoglobin borderline low
4239	Haemoglobin high
423Z.00	Haemoglobin estimation NOS
4232	Haemoglobin requested
4235	Haemoglobin low
4237	Haemoglobin normal
423..11	Hb estimation

Supplementary Table C.9. Read codes and terms for albumin level.

Read code	Read term
44M4.00	Serum albumin
44M4000	Serum albumin normal
44MI.00	Plasma albumin level
44M4100	Serum albumin low

Supplementary Table C.10. Read codes and terms for albumin level.

Read code	Read term
44CC000	C reactive protein normal
44CC100	C reactive protein abnormal
44CC.00	Plasma C reactive protein
44CS.00	Serum C reactive protein level

Supplementary Table C.11. Read codes and terms for breast cancer.

Read code	Read term
B830100	Intraductal carcinoma in situ of breast
Byu6.00	[X]Malignant neoplasm of breast
B34yz00	Malignant neoplasm of other site of female breast NOS
B342.00	Malignant neoplasm of upper-inner quadrant of female breast
BB91.11	[M]Duct carcinoma NOS
B34..00	Malignant neoplasm of female breast
B830.00	Carcinoma in situ of breast
B343.00	Malignant neoplasm of lower-inner quadrant of female breast
B341.00	Malignant neoplasm of central part of female breast
BB91100	[M]Infiltrating duct and lobular carcinoma
B340100	Malignant neoplasm of areola of female breast
B83..00	Carcinoma in situ of breast and genitourinary system
BB9E000	[M]Intraductal carcinoma and lobular carcinoma in situ
B34y000	Malignant neoplasm of ectopic site of female breast
B34y.00	Malignant neoplasm of other site of female breast
BB9K.00	[M]Paget's disease and infiltrating breast duct carcinoma

BB9J.00	[M]Paget's disease, mammary
B344.00	Malignant neoplasm of upper-outer quadrant of female breast
B340000	Malignant neoplasm of nipple of female breast
BA03.00	Neoplasm of unspecified nature of breast
ByuFG00	[X]Other carcinoma in situ of breast
B345.00	Malignant neoplasm of lower-outer quadrant of female breast
BB94.00	[M]Juvenile breast carcinoma
B3...11	Carcinoma of bone, connective tissue, skin and breast
B3y..00	Malig neop of bone, connective tissue, skin and breast OS
B830000	Lobular carcinoma in situ of breast
B346.00	Malignant neoplasm of axillary tail of female breast
BB9K000	[M]Paget's disease and intraductal carcinoma of breast
B340z00	Malignant neoplasm of nipple or areola of female breast NOS
B34z.00	Malignant neoplasm of female breast NOS
B347.00	Malignant neoplasm, overlapping lesion of breast
BB9E.00	[M]Lobular carcinoma in situ
B3z..00	Malig neop of bone, connective tissue, skin and breast NOS
BB9J.11	[M]Paget's disease, breast
BB90.00	[M]Intraductal carcinoma, noninfiltrating NOS
BB9F.00	[M]Lobular carcinoma NOS
B340.00	Malignant neoplasm of nipple and areola of female breast
B3...00	Malig neop of bone, connective tissue, skin and breast
BB91.00	[M]Infiltrating duct carcinoma
B34..11	Ca female breast
BB94.11	[M]Secretory breast carcinoma

Appendix D: Model specifications (RE Chapter 3)

This Appendix contains details of the logistic regression model specifications used in the analyses in **Chapter 3**. The Odds ratios for terms included in each model and the location of knots in each model are described in order to aid interpretation and enable study replication. In all models, log CA125 was centred on a value of 3 prior to analysis. Where age was included, it was centred on a value of 55.

Contents:

Supplementary Table D.10: *Odds ratios for terms included in the ≥50 years CA125 ovarian cancer model. Results displayed in Figure 3.6.*

Supplementary Table D.11: *Knots included in the age / CA125 ovarian cancer model. Results displayed in Figure 3.7.*

Supplementary Table D.12: *Odds ratios for terms included in the age / CA125 ovarian cancer model. Results displayed in Figure 3.7.*

Supplementary Table D.1. Knots included in the overall CA125 ovarian cancer model (women of all ages). Results displayed in Figure 3.5.

Knot	**Placement on centred log CA125, (Quantile)**
k1	-1.390562 (0.05)
k2	-0.8027754 (0.275)
k3	-0.5150933 (0.5)
k4	-0.1667867 (0.725)
k5	0.7376696 (0.95)

Supplementary Table D.2. Odds ratios for terms included in the overall CA125 ovarian cancer model (women of all ages). Results displayed in Figure 3.5.

Variable	**OR (95% CI)**
CA125 spline 1	1.93881 (0.1789538 - 21.00534)
CA125 spline 2	0.0105733 (6.00×10^{-09} - 18640.92)
CA125 spline 3	3.08×10^{27} (0.0000101 - 9.42×10^{59})
CA125 spline 4	7.66×10^{-50} (1.03×10^{-89} - 5.72×10^{-10})
Constant	*0.0018886 (0.0001076 - 0.033156)*

Supplementary Table D.3. Knots included in the overall CA125 all cancer model (women of all ages). Results displayed in Figure 3.5.

Knot	**Placement on centred log CA125, (Quantile)**
k1	-1.208241 (0.10)
k2	-0.5150933 (0.5)
k3	0.3322045 (0.90)

Supplementary Table D.4. Odds ratios for terms included in the overall CA125 all cancer model (women of all ages). Results displayed in Figure 3.5.

Variable	**OR (95% CI)**
CA125 spline 1	1.506687 (1.205222 - 1.883559)
CA125 spline 2	1.774489 (1.472288 - 2.138719)
Constant	0.0239013 (0.0201886 - 0.0282968)

Supplementary Table D.5. Knots included in the overall CA125 invasive ovarian cancer model (women of all ages). Results displayed in Figure 3.5.

Knot	Placement on centred log CA125, (Quantile)
k1	-1.390562 (0.05)
k2	-0.8027754 (0.275)
k3	-0.5150933 (0.5)
k4	-0.1667867 (0.725)
k5	0.7376696 (0.95)

Supplementary Table D.6. Odds ratios for terms included in the overall CA125 invasive ovarian cancer model (women of all ages). Results displayed in Figure 3.5.

Variable	OR (95% CI)
CA125 spline 1	0.918744 (0.085896 - 9.826892)
CA125 spline 2	0.0015715 ($8.94x10^{-11}$ - 27620.28)
CA125 spline 3	$2.05x10^{40}$ (0.9907357 - $4.23x10^{80}$)
CA125 spline 4	$1.74x10^{-72}$ ($4.8x10^{-124}$ - $6.36x10^{-21}$)
Constant	*0.0005503*

Supplementary Table D.7. Knots included in the <50 years CA125 ovarian cancer model. Results displayed in Figure 3.6.

Knot	Placement on centred log CA125, (Quantile)
k1	-1.295252 (0.05)
k2	-0.6021047 (0.35)
k3	-0.1667867 (0.65)
k4	0.7358818 (0.95)

Supplementary Table D.8. Odds ratios for terms included in the <50 years CA125 ovarian cancer model. Results displayed in Figure 3.6.

Variable	OR (95% CI)
CA125 spline 1	0.7911065 (0.0342347 - 18.28112)
CA125 spline 2	801.5433 (0.2558281 - 2511342)
CA125 spline 3	$1.37x10^{-10}$ ($1.86x10^{-21}$ - 10.16702)
Constant	*0.0003367 (0.0000169 - 0.0066908)*

Supplementary Table D.9. Knots included in the ≥50 years CA125 ovarian cancer model. Results displayed in Figure 3.6.

Knot	Placement on centred log CA125, (Quantile)
k1	-1.410765 (0.05)
k2	-0.9205585 (0.275)
k3	-0.6021047 (0.5)
k4	-0.2274113 (0.725)
k5	0.7376696 (0.95)

Supplementary Table D.10. Odds ratios for terms included in the ≥50 years CA125 ovarian cancer model. Results displayed in Figure 3.6.

Variable	OR (95% CI)
CA125 spline 1	1.930133 (0.1115593 - 33.39403)
CA125 spline 2	0.0000447 $4.05x10^{-15}$ - 492193.3)
CA125 spline 3	$3.49x10^{34}$ (0.000021 - $5.80x10^{73}$)
CA125 spline 4	$3.43x10^{-57}$ ($1.2x10^{-100}$ - $9.73x10^{-14}$)
Constant	0.0023171 (0.0000615 - 0.0872512)

Supplementary Table D.11. Knots included in the age / CA125 ovarian cancer model. Results displayed in Figure 3.7.

Knot	Placement on centred age and centred log CA125, (Quantile)
Age k1	-24 (0.05)
Age k2	-9 (0.275)
Age k3	-1 (0.5)
Age k4	10 (0.725)
Age k5	27 (0.95)
CA125 k1	-1.390562 (0.05)
CA125 k2	-0.8027754 (0.275)
CA125 k3	-0.5150933 (0.5)
CA125 k4	-0.1667867 (0.725)
CA125 k5	0.7376696 (0.95)

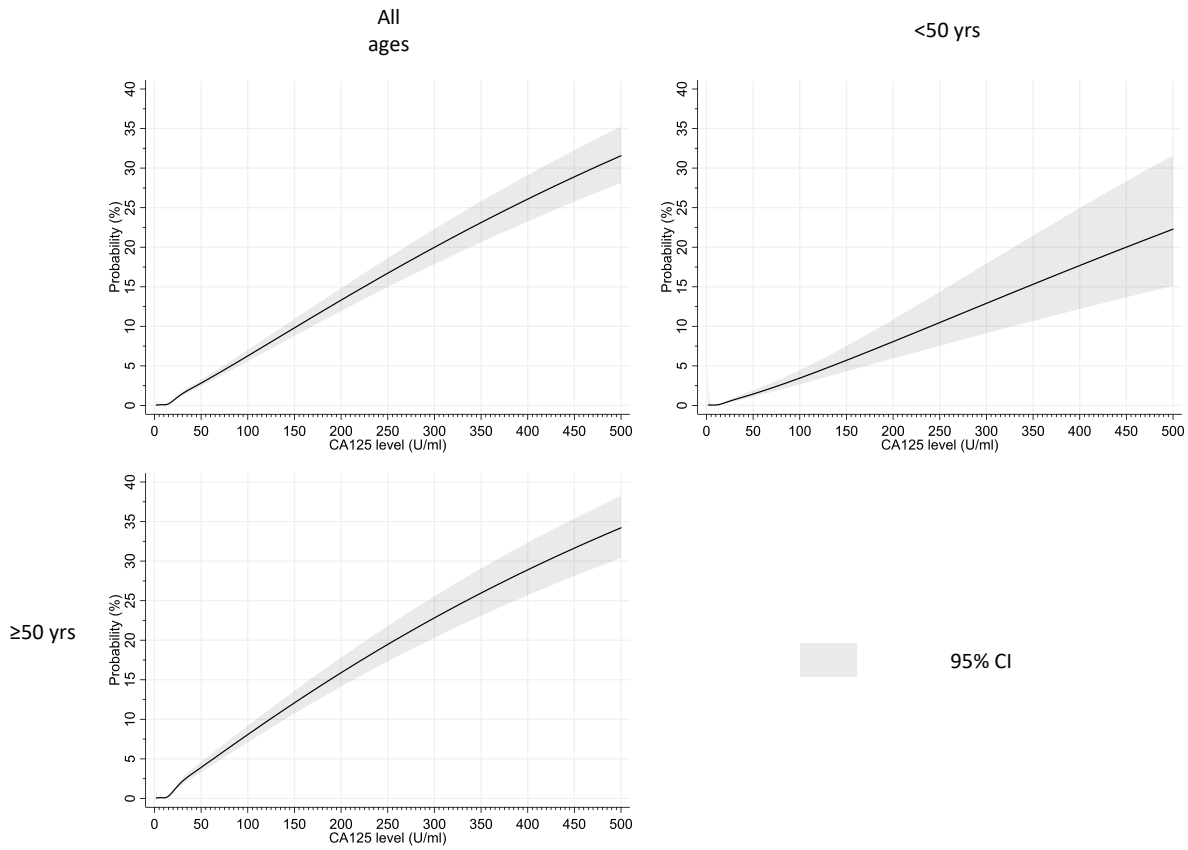
Supplementary Table D.12. Odds ratios for terms included in the age / CA125 ovarian cancer model. Results displayed in Figure 3.7.

Variable	OR (95% CI)
Age spline 1	0.9263145 (0.8896194 - 0.9645232)
Age spline 2	1.746288 (1.364622 - 2.234702)
Age spline 3	0.1047997 (0.0284088 - 0.3866058)
Age spline 4	7.06957 (1.294418 - 38.61104)
CA125 spline 1	2.26625 (0.2010149 - 25.54979)
CA125 spline 2	0.0046567 ($2.28x10^{-09}$ - 9515.724)
CA125 spline 3	$1.25x10^{31}$ (0.0223821 - $6.99x10^{63}$)
CA125 spline 4	$3.40x10^{-56}$ ($2.33x10^{-96}$ - $4.96x10^{-16}$)
Constant	*0.0002362 (0.0000114 - 0.0048961)*

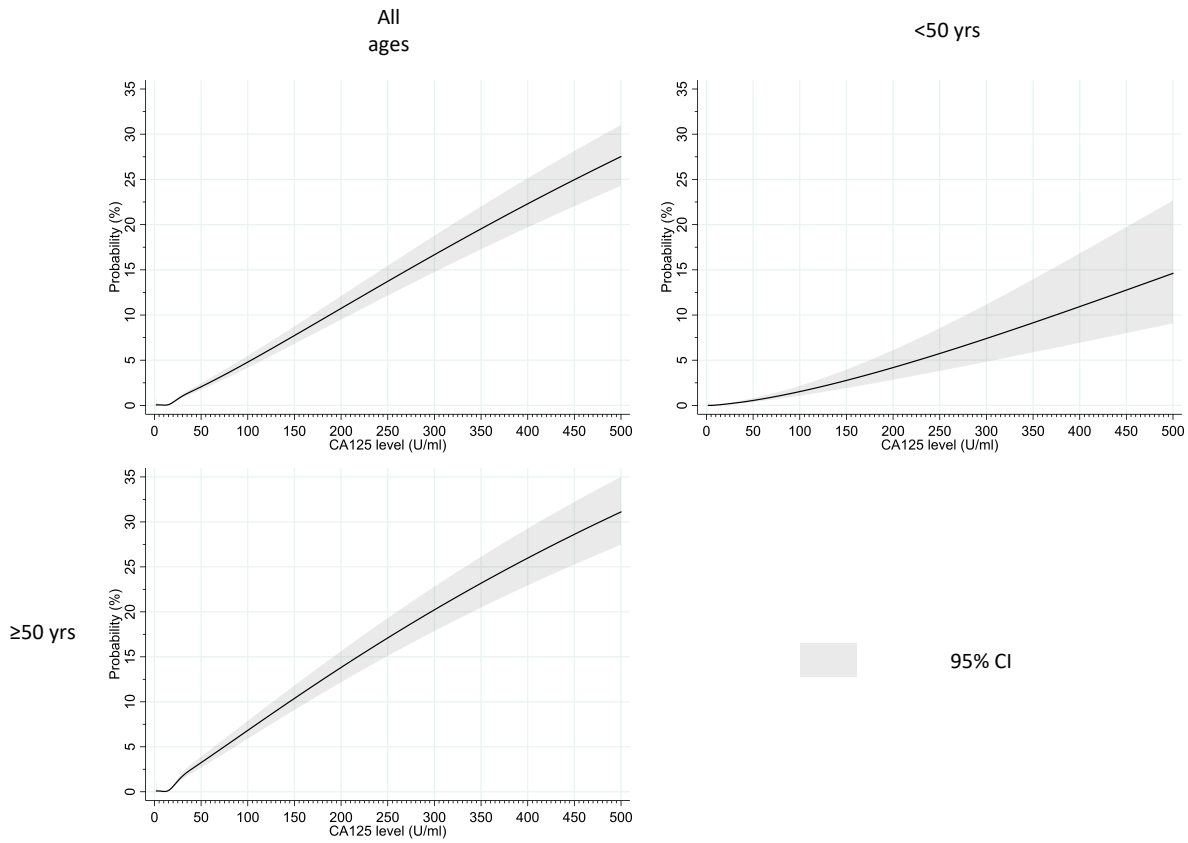
Appendix E: Supplementary figures (RE Chapter 3)

This Appendix comprises supplementary figures relating to Chapter 3.

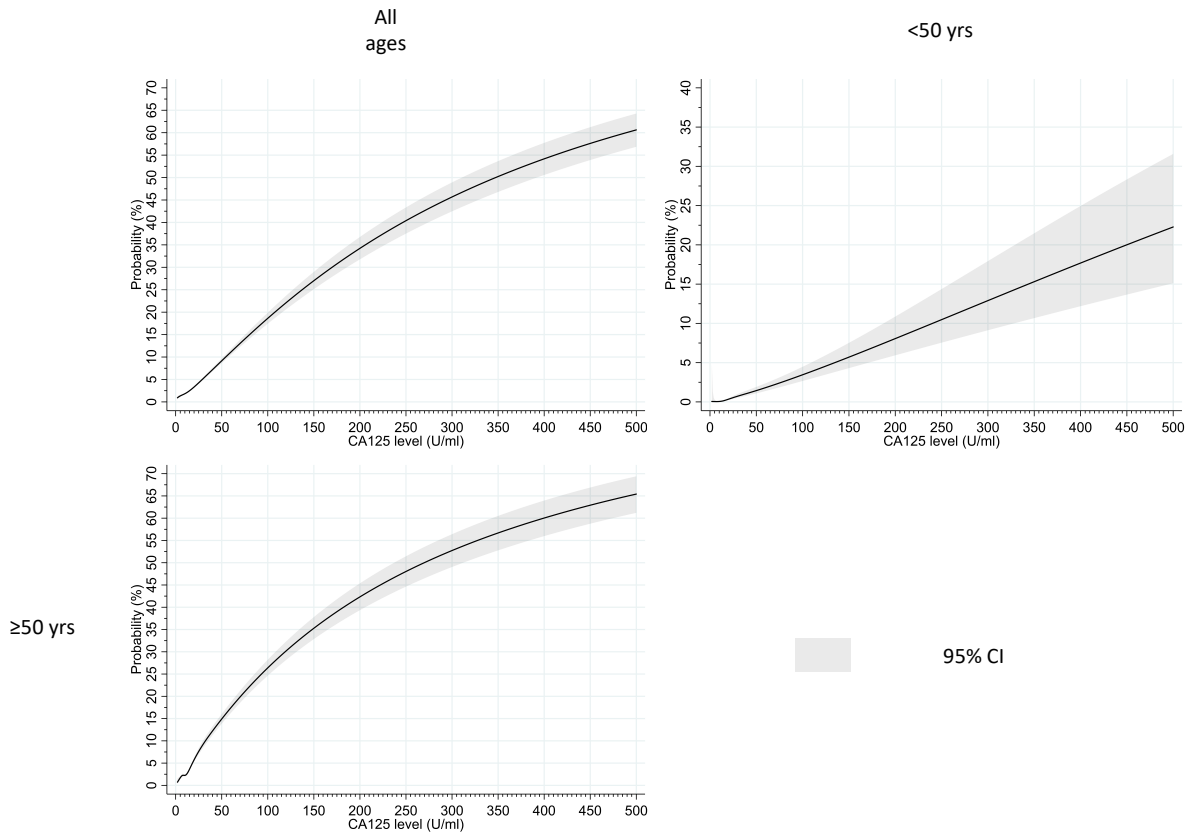
Contents:



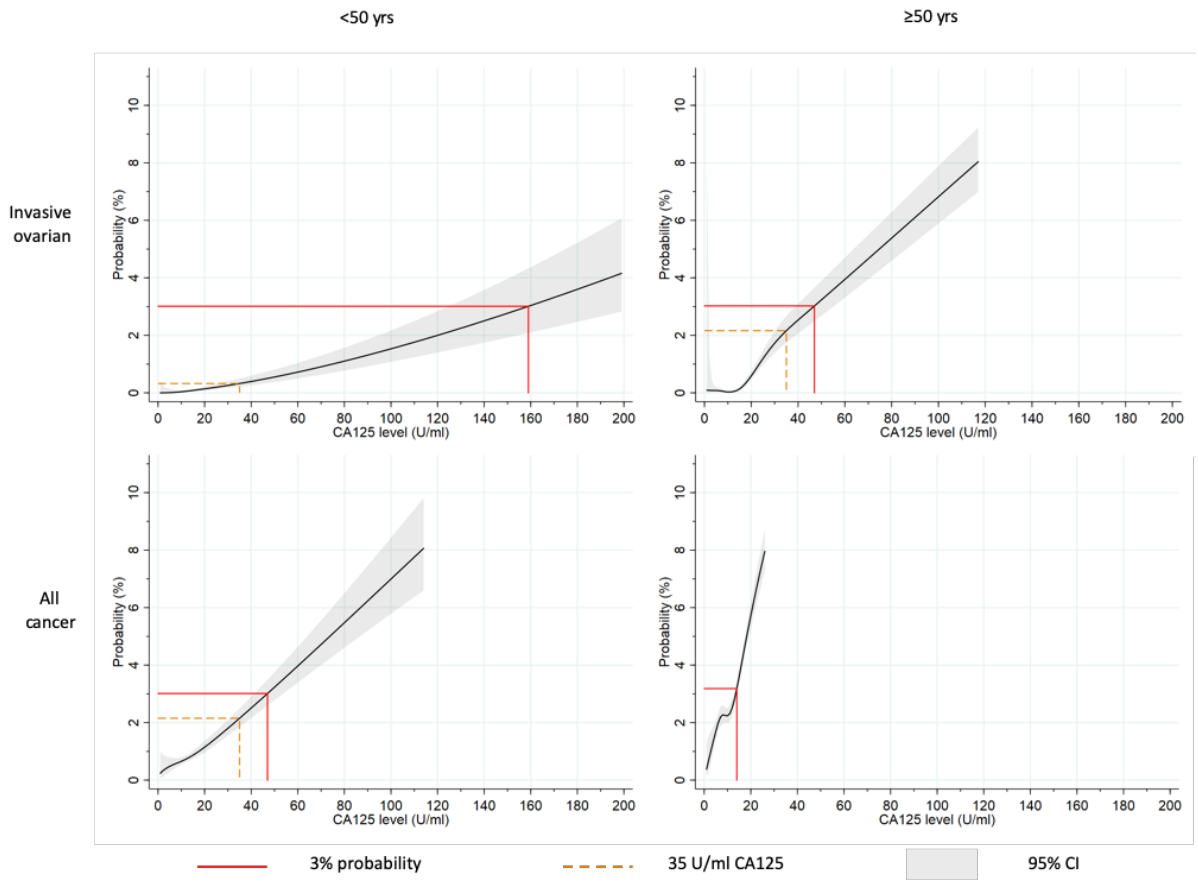
Supplementary Figure E.1. Estimated probabilities for ovarian cancer up to CA125 500 U/ml. CI= Confidence Interval.



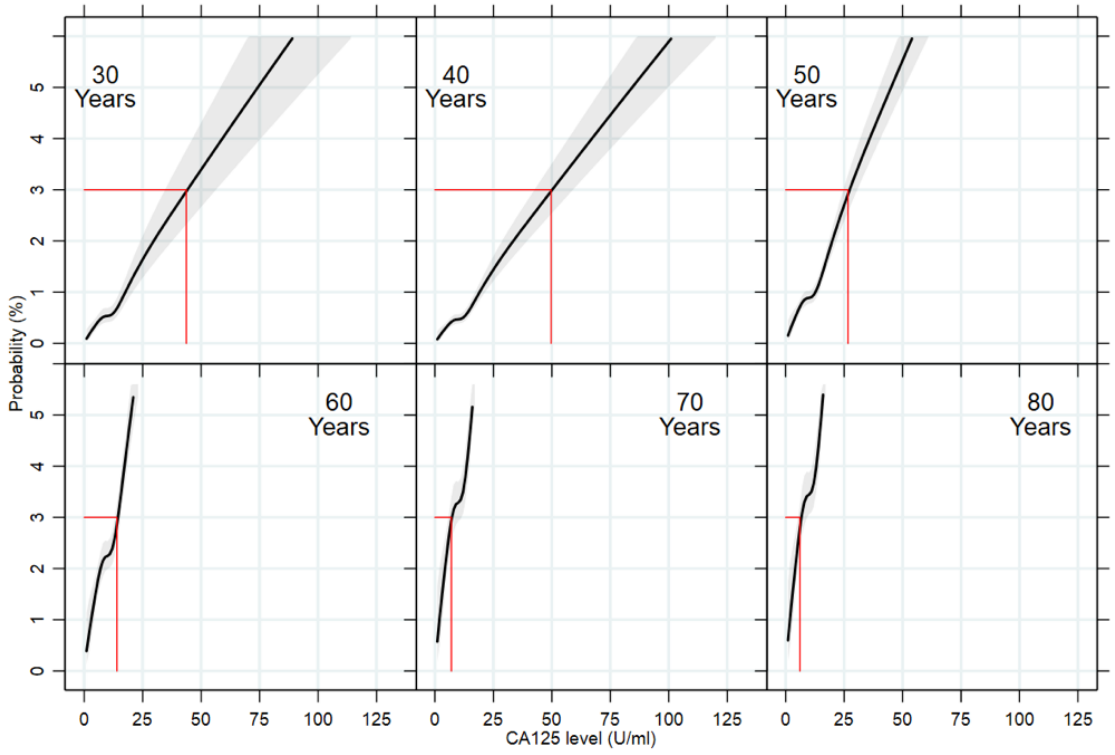
Supplementary Figure E.2. Estimated probabilities for invasive ovarian cancer up to CA125 500 U/ml. CI= Confidence Interval.



Supplementary Figure E.3. Estimated probabilities for all cancers up to CA125 500 U/ml.
CI= Confidence Interval.



Supplementary Figure E.4. Relationship between CA125 level and estimated probability of invasive ovarian cancer and all cancers in women <50 years and ≥50 years of age.



Supplementary Figure E.5. Relationship between CA125 level and estimated probability of all cancers for women of different ages.

Probabilities are shown in relation to CA125 level for women of 30, 40, 50, 60, 70 and 80 years of age. CA125 levels which correspond to the closest integer probabilities of 3% are indicated in red. 95% confidence intervals are displayed.

Prof Gary Abel provided Stata code which was used to lay out this figure.

Appendix F: Supplementary tables (RE Chapter 4)

This Appendix comprises supplementary results relating to Chapter 4.

Contents:

Supplementary Table F.1. Baseline characteristics of women with a recorded stage for ovarian cancer.

	n	Mean age at diagnosis [range]	Patients with a symptom of possible ovarian cancer recorded pre-testing, n (%)*	Townsend score, n (%)*
Abnormal CA125	304	64 [22-88]	189 (62.2)	Level 1: 71 (23.4) Level 2: 85 (28) Level 3: 66 (21.7) Level 4: 55 (18.1) Level 5: 27 (8.9)
Normal CA125	77	59 [20-87]	44 (57.1)	Level 1: 15 (19.5) Level 2: 23 (29.9) Level 3: 19 (24.7) Level 4: 11 (14.3) Level 5: 9 (11.7)
Overall cohort	381	63 [20-88]	233 (61.2)	Level 1: 86 (22.6) Level 2: 108 (28.4) Level 3: 85 (22.3) Level 4: 66 (17.3) Level 5: 36 (9.5)

*Percentage of each group with symptoms and Townsend score

Supplementary Table F.2. Association between CA125 test result, age, presence/absence of a recorded symptom and Townsend score with early stage (I-II) at diagnosis for invasive tumours.

	Unadjusted		Adjusted	
	OR	**95% CI (p value)**	**OR**	**95% CI (p value)**
Abnormal CA125	Reference	-	Reference	-
Normal CA125	8.6	4.1-18.0 (<0.001)	9.1	4.1-20.2 (<0.001)
Age (years)	0.95	0.93-0.97 (<0.001)	0.94	0.92-0.97 (<0.001)
No symptom record	Reference	-	-	-
Symptom record	0.50	0.31-0.81 (0.005)	0.35	0.19-0.63 (<0.001)
Townsend score	-	(0.39)*	-	(0.77)*
Townsend 1	Reference	-	-	-
Townsend 2	1.9	0.95-3.6 (0.07)	1.5	0.70-3.4 (0.28)
Townsend 3	1.3	0.64-2.7 (0.46)	1.1	0.47-2.5 (0.85)
Townsend 4	1.4	0.66-3.0 (0.37)	1.3	0.53-3.0 (0.60)
Townsend 5	0.92	0.32-2.6 (0.87)	0.86	0.26-2.8 (0.80)

*P value for effect of Townsend score as a whole.

Appendix G: PRISMA checklist (RE Chapter 5)

This appendix comprises the PRISMA checklist completed for the systematic review presented in **Chapter 5**.

Section/topic	#	Checklist item	Location
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Chapter Title
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	N/A (no study abstract)
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Section 5.1
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Section 5.1
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Section 5.2.2
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Sections 5.2.3 and 5.2.4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Section 5.2.3
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix H

Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Section 5.2.5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Section 5.2.6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Section 5.2.6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Section 5.2.8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Section 5.2.7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Section 5.2.7
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Section 5.3.1 and Fig 5.1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Section 5.3.2 and table 5.1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Section 5.3.3 and Figure 5.2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Sections 5.3.4-5.3.6
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-	N/A

		regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Section 5.4.1
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	5.4.2
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	5.5
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Acknowledgements section of thesis

Appendix H: Supplementrary inforation (RE Chapter 5)

Supplementary Box H.1: *MEDLINE search strategy.*

Supplementary Table H.1: *Specific study exclusions.*

Supplementary Table H.2: *Tool specifications.*

Supplementary Table H.3: *Deviations from the original Goff SI in validation studies.*

1. ((ovar* or fallopian or peritone*) and (cancer* or neoplas* or tumour* or tumor* or malignan*)).ti,ab.
2. exp Ovarian Neoplasms/ or exp Fallopian Tube Neoplasms/ or exp Peritoneal Neoplasms/
3. 1 or 2
4. symptom*.ti,ab.
5. exp symptom assessment/
6. 4 or 5
7. (risk* or probabilit* or likelihood* or chance*).ti,ab.
8. exp Risk/ or exp Risk factors/ or exp Probability/
9. predict*.ti,ab.
10. exp "Early Detection of Cancer"/
11. diagnos*.ti,ab.
12. 7 or 8 or 9 or 10 or 11 (5845128)
13. (model* or algorithm* or tool* or index* or score* or rule*).ti,ab.
14. exp models, statistical/ or exp algorithms/
15. 13 or 14
16. 3 and 6 and 12 and 15
17. limit 16 to yr="2000 -Current"

Supplementary Box H.1. MEDLINE search strategy.

Supplementary Table H.1. Specific study exclusions.

Author, Date	Specific exclusions
Lurie, 2009	*Controls*: Hx OC, no intact ovaries
Rossing, 2010	Non-English speakers, no residential telephone *Controls*: Hx OC, no intact ovaries
Jordan, 2010	*Cases:* language difficulties, mental incapacity, illness *Controls:* language difficulties, illness, previous ovarian cancer or previous bilateral oophorectomy [252]
Hamilton, 2009	No entry in the records ≤1 year pre-diagnosis (cases), previous OC or bilateral oophorectomy, lived outside study area at time of diagnosis (cases) [91]
Hippisley-Cox, 2012	Hx bilateral oophorectomy or OC, 'red flag symptom' ≤12 months before study entry date[a], no postcode related Townsend score
Hippisley-Cox, 2013	'Red flag symptom' ≤12 months before study entry date, no postcode-related Townsend score
Grewal, 2013, UK	No entry in the records ≤1 year pre-diagnosis (cases), previous OC or bilateral oophorectomy, lived outside study area at time of diagnosis (cases) [91]
Collins, 2013	As per Hippisley-Cox, 2011
Goff, 2006	Screening control criteria as outlined in OCEDS [254]
Anderson, 2008	Screening control criteria as outlined in OCEDS [254]
Anderson, 2010	Screening control criteria as outlined in OCEDS [254b]

Lim, 2012	*Controls:* Hx of bilateral oophorectomy or OC, active malignancy, increased risk of familial OC, not post menopause (as per UKCTOCS trial criteria) [154]
Kim, 2009	*Pap smear controls:* Hx of gynaecological malignancy, no intact ovaries or uterus
Macuks, 2011	Severe co-morbidities, previous or other coexisting malignancies
Shetty, 2015	*Controls (gynae check-up group):* no ovaries, no intact uterus
Jain, 2018	*Cases:* Hx of ovarian cancer, Hx bilateral oophorectomy, recall difficulty, inoperable

[a] Loss of appetite, weight loss, abdominal pain, abdominal distension, rectal bleeding, or postmenopausal bleeding

[b] Patients with known BRCA mutations excluded during study

Supplementary Table H.2. Tool specifications.

Tool	First study (author, year)	Specification
Symptom checklists		
Goff SI	Goff, 2007	Tool positive if any of pelvic/abdominal pain, increased abdominal size/bloating, and difficulty eating/feeling full occurred >12 times per month but were present for <1 year
Modified Goff SI 1	Kim, 2009	Tool positive if any of pelvic/abdominal pain, urinary urgency/frequency, increased abdominal size/bloating, difficulty eating/feeling full present for <1 year that occurred >12 times per month
Lurie 7-SI	Lurie, 2009	Tool positive if any of distended abdomen (defined as "persistent distended and hard abdomen"), abnormal vaginal bleeding (defined as "vaginal bleeding not associated with periods"), palpable abdominal mass (defined as "a palpable abdominal mass that woman herself had noticed"), abdominal pain (defined as "persistent abdominal or pelvic pain or discomfort"), urinary symptoms (defined as "urinary frequency, difficulty emptying urinary bladder, or dysuria"), bowel symptoms (defined as "unusual bowel irregularity such as diarrhoea or constipation, flatulence, or bloating"), and fatigue/appetite loss (defined as "persistent fatigue or loss of appetite"), present in previous 12 months
Lurie 5-SI	Lurie, 2009	Tool positive if any of distended abdomen (defined as "persistent distended and hard abdomen"), abnormal vaginal bleeding (defined as "vaginal bleeding not associated with periods"), palpable abdominal mass (defined as "a palpable abdominal mass that woman herself had noticed"), abdominal pain (defined as "persistent abdominal or pelvic pain or discomfort"), and urinary symptoms (defined as "urinary frequency, difficulty emptying urinary bladder, or dysuria"), present in previous 12 months
Lurie 4-SI	Lurie, 2009	Tool positive if any of distended abdomen (defined as "persistent distended and hard abdomen"), abnormal vaginal bleeding (defined as "vaginal bleeding not associated with periods"), palpable abdominal mass (defined as "a palpable abdominal mass that woman herself had noticed"), and abdominal pain (defined as "persistent abdominal or pelvic pain or discomfort"), present in previous 12 months

Lurie 3-SI	Lurie, 2009	Tool positive if any of distended abdomen (defined as "persistent distended and hard abdomen"), abnormal vaginal bleeding (defined as "vaginal bleeding not associated with periods"), and palpable abdominal mass (defined as "a palpable abdominal mass that woman herself had noticed"), present in previous 12 months
Hamilton SI	Hamilton, 2009	Tool positive if any of bloating, urinary frequency, rectal bleeding, postmenopausal bleeding, loss of appetite, abdominal pain, abdominal distension, present in the previous 12 months
SGO consensus criteria	Rossing, 2010	Tool positive if any of bloating or feeling full, pelvic or abdominal pain, or urinary urgency or frequency, present for at least 1 month, with an onset of less than 12 months
Lim SI 1	Lim, 2012	Tool positive if any of pelvic/abdominal pain or discomfort, loss of appetite or feeling full quickly, weight loss, increase in abdominal size, abdomen feels bloated, and able to feel a lump in the abdomen in previous 12 months
Lim SI 2	Lim, 2012	Tool positive if any of pelvic abdominal pain or discomfort, loss of appetite, increase in abdominal size, able to feel a lump in the abdomen, and vaginal discharge in previous 12 months
Hippisley-Cox SI	Hippisley-Cox, 2012	Tool positive if currently consulting general practitioner with first onset of any of abdominal pain, abdominal distension, appetite loss, rectal bleeding, postmenopausal bleeding, weight loss.
Modified Goff SI 2	Shetty, 2015	Tool positive if any of abdominal/pelvic pain, increased abdominal size/bloating, difficulty in eating/feeling full and urinary frequency/urgency, loss of appetite/weight, occurred >12 times per month and time since onset was <1 year
Augmented symptom checklist		
Goff SI + CA125	Anderson, 2008	The threshold for a positive CA125 test was determined by dichotomizing CA 125 at the 95th percentile in the control group (threshold approx. 30 u/ml).
Goff SI + HE4	Anderson, 2010	The threshold for a positive HE4 test was determined by dichotomizing HE4 at the 95th percentile in the control group.
Goff SI + HE4 + CA125	Anderson, 2010	The threshold for a positive HE4 and CA125 tests were determined by dichotomizing HE4 at the 95th percentile in the control group. Study evaluated several thresholds for a positive tool (**Table 4).**
Goff SI + CA125 + menopause	Macuks, 2011	CA125 thresholds: 25 U/ml, 35 U/ml and 65 U/ml examined. Definition of menopause not specified.
Prediction models		
QCancer Ovarian	Hippisley-Cox, 2012	The prediction model included age, family history of ovarian cancer, haemoglobin <110 g/L in past year, currently consulting general practitioner with first onset of any of abdominal pain, abdominal distension, appetite loss, rectal bleeding, postmenopausal bleeding, weight loss. Tool threshold was set based on risk level e.g. 10% of women at highest risk deemed tool positive.
QCancer Female	Hippisley-Cox, 2013	The prediction model included age, BMI, Townsend score, smoking status, alcohol status, family history of gastrointestinal cancer, family history of breast cancer, family history of ovarian cancer, type 2 diabetes, COPD, endometrial hyperplasia or polyp, chronic pancreatitis. Current: loss of appetite, unintentional weight loss, abdominal pain, abdominal swelling, difficulty swallowing, heartburn or indigestion, rectal bleeding,

		blood in urine, blood in vomit, blood when cough, postmenopausal bleeding, irregular menstrual bleeding, vaginal bleeding after sex, a breast lump, breast skin tethering or nipple discharge, breast pain, a lump in your neck, night sweats, a venous thromboembolism. In the last year seen GP with: change in bowel habit, constipation, cough, unexplained bruising, anaemia (haemoglobin <11g/dL). Tool threshold was set based on risk level e.g. 10% of women at highest risk deemed tool positive.
OC Scores A	Grewal, 2013	Variables: bloating, urinary frequency, rectal bleeding, postmenopausal bleeding, loss of appetite, abdominal pain, abdominal distension, present in the previous 12 months. Model used conditional logistic logarithmic odds ratio of each symptom, to three significant figures. Various threshold reported (**Table 4**).
OC Scores B	Grewal, 2013	Variables: bloating, urinary frequency, rectal bleeding, postmenopausal bleeding, loss of appetite, abdominal pain, abdominal distension, present in the previous 12 months. Model used the conditional logarithmic odds ratio of each variable rounded to the nearest integer. Various threshold reported (**Table 4**).
OC Scores C	Grewal, 2013	Variables: Age (≥50 years / < 50 years), bloating, urinary frequency, rectal bleeding, postmenopausal bleeding, loss of appetite, abdominal pain, abdominal distension, present in the previous 12 months. Various threshold reported (**Table 4**).

Supplementary Table H.3. Deviations from the original Goff SI in validation studies.

Study	Deviation
Rossing, 2010	1) Symptom criteria listed as "bloating or feeling full", whereas original Goff SI includes "Increased abdominal size/bloating" and "difficulty eating/feeling full". 2) Duration/frequency of symptoms criteria was "present at least daily for at least 1 week", whereas the original Goff SI criteria is >12x/month.
Jordan, 2010	Duration/frequency of symptoms criteria was ">2 weeks in previous 12 months" whereas the original Goff SI criteria is >12x/month.
Lim, 2012	Duration/frequency of symptoms criteria was "occurred 16-31 days per month" for interview and questionnaire study components, and no frequency criteria was applied in the GP notes study component. The original Goff SI criteria is >12x/month.

Appendix I: Excluded variables (RE Chapter 6).

This appendix lists the variables considered but ultimately not taken forwards into data driven selection procedures in **Chapter 6**. The rationale for exclusion is also detailed for each variable.

Variable	Rationale for exclusion			Elaboration
	Limited evidence	Recording / coding concerns	Similar variable included	
Blood biomarkers				
HE4		•		Identified in the ovarian cancer tool review but not in routine use in GP (no Read code).
Erythrocyte Sedimentation Rate (ESR) and Plasma Viscosity (PV)			•	Recognised predictors for cancer in primary care, unknown association with ovarian cancer.[175] CRP included in preference to ESR and PV as more common in the baseline cohort.
Risk / protective factors				
Family history of ovarian and breast cancer		•		Risk of ovarian cancer is greater in women with family history of breast or ovarian cancers.[25,26] However, preliminary searches in the dataset showed that the prevalence of relevant codes for family history of these cancers in the CPRD is much lower in our cohort than in published studies - likely significant missing data. Potential for bias.
Known germline BRCA (1 or 2) mutation		•		BRCA mutations and family history of BRCA associated cancers are a strong risk factor for ovarian cancer.[25] Preliminary searches within CPRD records of the baseline cohort showed that these codes were not recorded in the cohort.
Asbestos exposure		•		Evidence from systematic reviews and meta-analyses indicates that asbestos exposure is associated

				with an increased risk of ovarian cancer.[39,41] *Consensus group* - concerns over likely missing data pertaining to asbestos exposure in GP records.
Breastfeeding	●	●		Evidence from systematic reviews and meta-analyses indicate that breastfeeding is a protective factor for ovarian cancer.[25,274,316] However, 2018 World Cancer Research Fund report only found "limited- suggestive evidence" for an association.[274] *Consensus group* - concerns over likely missing data (particularly historically) pertaining to breastfeeding in the CPRD.
Parity		●		There is evidence that increasing parity is associated with reduced ovarian cancer risk.[25,274,317,318] *Consensus group* - concerns over recording of historical births in the CPRD.
Early menarche		●		There is evidence that early menarche is associated with an increased risk of ovarian cancer.[25,28,274] But, date of menarche is unlikely to be routinely recorded by GPs (<13,000 entries relating to menarche in the whole of the CPRD GOLD) - likely significant missing data. Practical issues foreseen by *consensus group* with patient recall if data entered manually into a tool during a GP consultation.
Late menopause		●		There is evidence that late menopause is associated with an increased risk of ovarian cancer.[25,29,274] *Consensus group* - concerns over likely missing data pertaining to menopause in the CPRD. Also, concern over possible inaccuracy of date of menopause even if codes recorded in the CPRD.
Diet / nutrition				Multiple dietary factors have been implicated in ovarian cancer

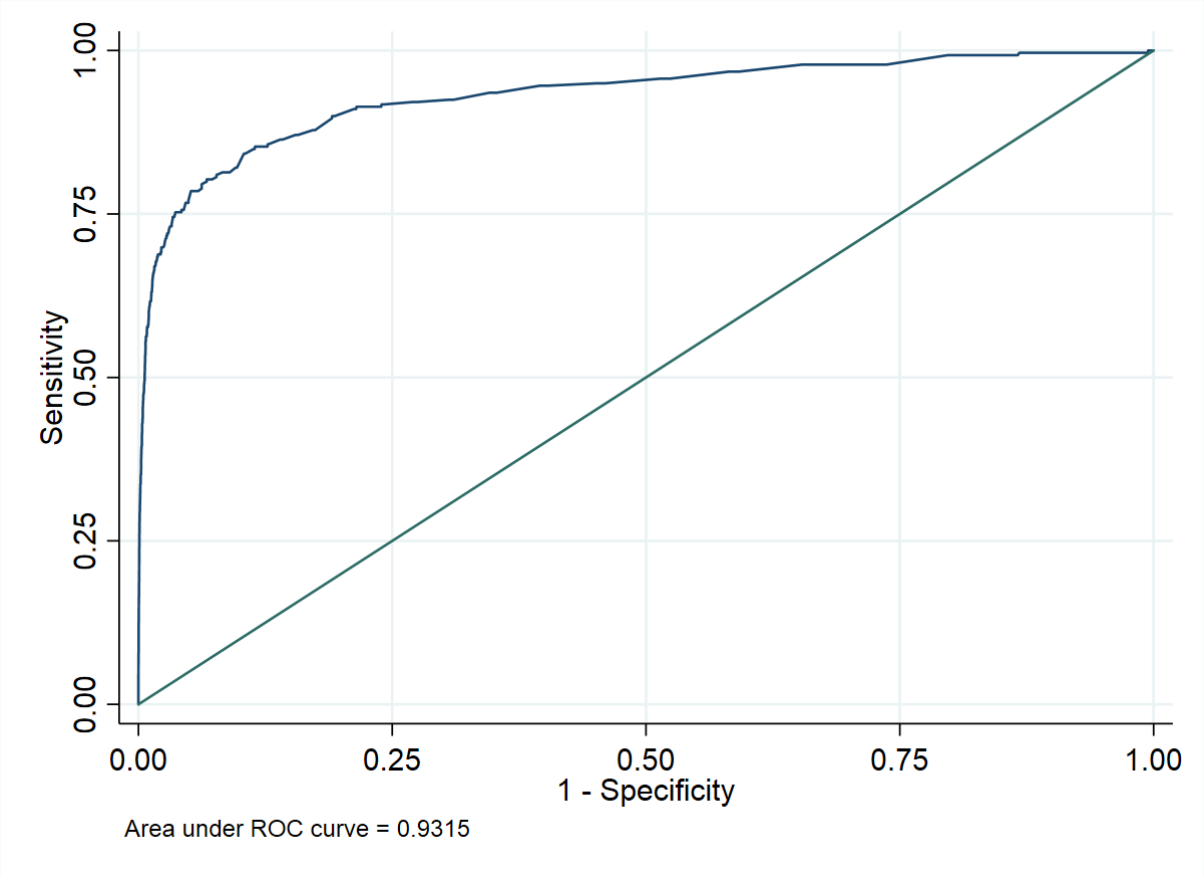
	●	●		development/risk.[319] However, the World Cancer Research Fund Report did not find enough evidence to draw conclusions on any dietary factor.[274] *Consensus group* - data on diet not routinely recorded in general practice records.
Sedentary lifestyle / physical activity	●	●		Some studies indicate that a sedentary lifestyle is associated with increased ovarian cancer risk. But, a meta-analysis of prospective studies in 2014 did not support that,[320] and insufficient evidence to draw a conclusion in World Cancer Research Fund report.[274] *Consensus group* - data not routinely collected by GPs and likely to be highly inconsistent.
Talc exposure	●	●		Listed as a risk factor for ovarian cancer by IARC but a recent meta-analysis did not find evidence that talc exposure increased the risk of ovarian cancer. [41,42] *Consensus group* - data not routinely recorded by GPs/likely bias.
Deprivation	●			Considered by models identified in the ovarian cancer tool review (Chapter 5), but not included in any of the final models. No studies identified which report an association between deprivation and ovarian cancer risk.
Hysterectomy (conservation of ovaries)	●			Older reports indicated that hysterectomy was a protective factor for ovarian cancer.[25,319] Multiple recent large studies found no evidence that hysterectomy reduces the overall risk of ovarian cancer.[321,322] Inconsistent evidence.
Pelvic inflammatory disease (PID)	●			Some studies have found an association between PID and ovarian cancer but the evidence has been inconsistent and contradictory.[319] A pooled analysis of 13 studies found no significant association between PID and

				ovarian cancer risk.[323] Inconsistent evidence.
Polycystic ovarian syndrome (PCOS)	●			Previously thought to increase ovarian cancer risk. Recent prospective studies and a meta-analysis found no significant increase in the risk of ovarian cancer in PCOS.[324,325]
Smoking status	●	●		Smoking is associated with an increased risk of the mucinous subtype of ovarian cancer.[40] Smoking is classified as an ovarian cancer carcinogenic agent by IARC.[41] However, no association between smoking and ovarian cancer overall was identified in a large pooled analysis study.[40] Smoking was not included in the final QCancer prediction model for ovarian cancer.[54] *Consensus group* - some concern about the accuracy of smoking status data in CPRD.
Alcohol	●			While an association was once postulated, recent studies, including a large pooled analysis, have found no association between alcohol and risk of ovarian cancer.[326]
Diabetes (1 and 2)	●			Some large population-based studies have not found any significant association while other studies and meta-analyses report an increased risk of ovarian cancer in women with diabetes.[327–330] Diabetes was dropped from final EPIC model (predicts future risk of ovarian cancer) as not significant.[331] *Consensus group - inconsistent evidence.*
Non-steroidal anti-inflammatory drugs (NSAIDs)	●	●		Some research indicates that NSAID use is inversely proportional to ovarian cancer risk.[25,332] However, two meta-analyses (2012, 2013) did not demonstrate a significant association between NSAID use and ovarian cancer risk.[333,334] NSAIDs are not noted as

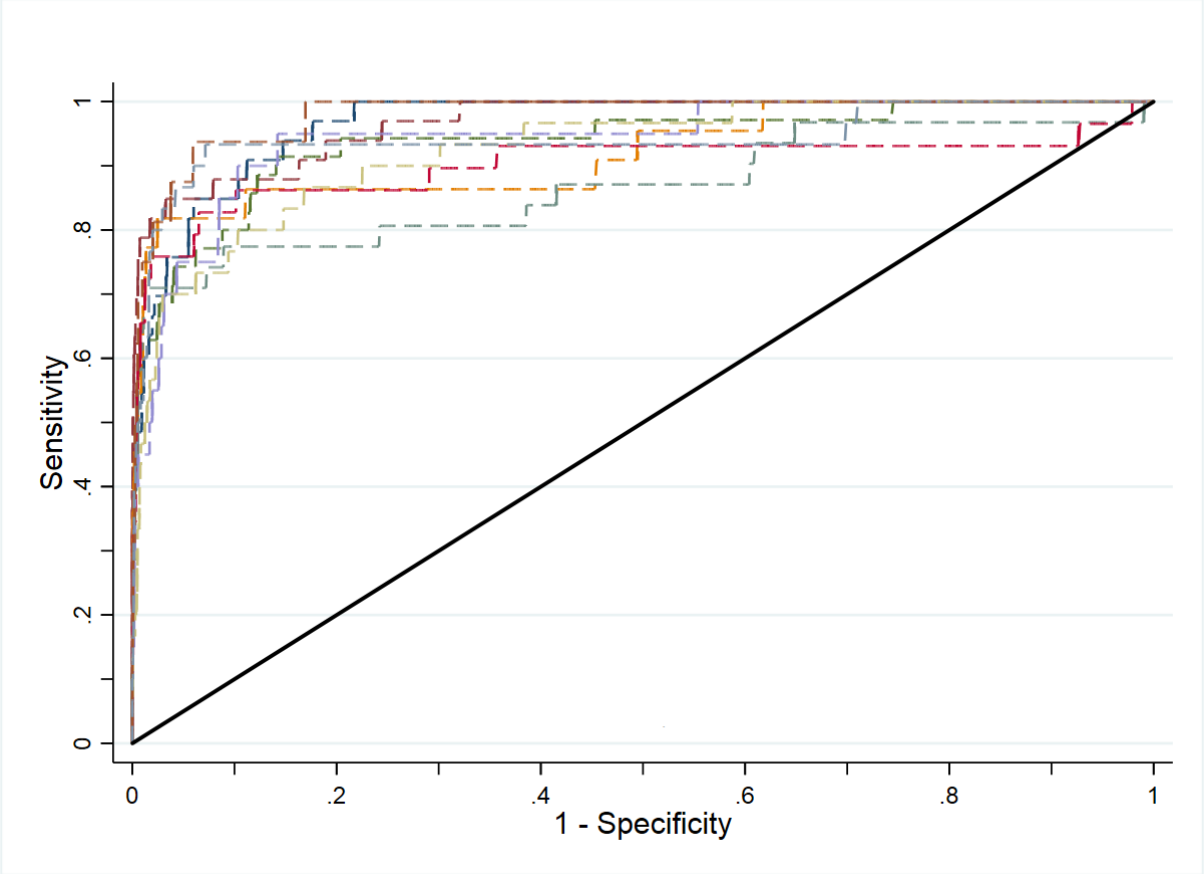
				a protective factor for ovarian cancer by IARC.[41] In adition, NSAIDs can be bought over the counter so CPRD is unlikely to capture the true scope of NSAID use.
Combined oral contraceptive (COC)		●		There is evidence from systematic reviews and meta-analyses that COC use is associated with lower risk of ovarian cancer.[25,30,274,335] Importantly, risk reduction is strongly related to duration of use.[30] Examination of the data in my cohort indicated that approaches to determine duration of use likely to be inaccurate due to limited information on number of tablets, prescription use and length. In addition, COC can be obtained from other locations e.g. sexual health centres, so GP records unlikely to be complete.
Hormone replacement therapy (HRT)		●		There is evidence from large prospective studies, systematic reviews and meta-analyses that HRT use is associated with increased ovarian cancer risk.[25,33,274,336] Duration of use is important so HRT was excluded (see COC).

Appendix J. ROC Curves (RE Chapter 6)

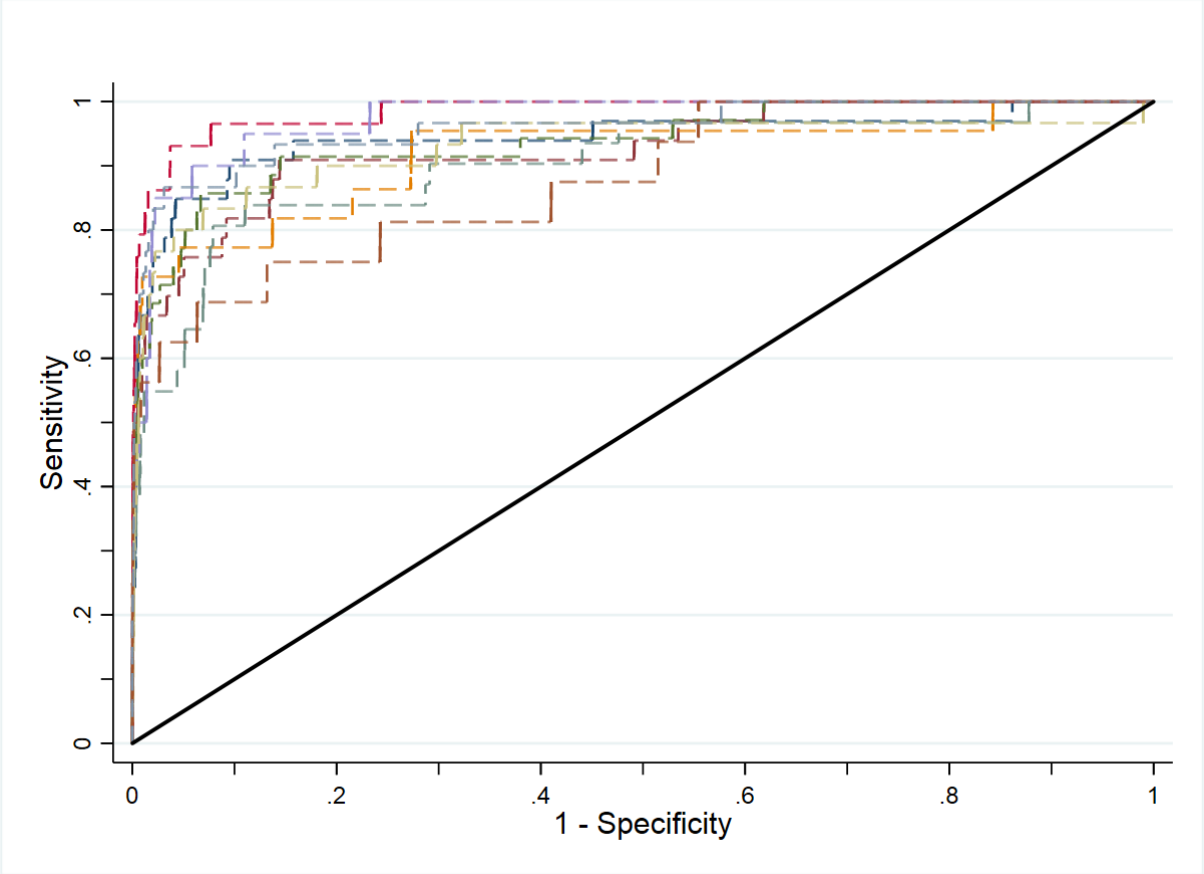
Contents:



Supplementary Figure J.1. Receiver operating characteristic curve for CA125.



Supplementary Figure J.2. Tenfold cross-validation receiver operating characteristic curve for Model 1.



Supplementary Figure J.3. Tenfold cross-validation receiver operating characteristic curve for Model 2, prepared using imputation set 20 as an example.

Note: To calculate the overall cross-validation AUC for Model 2, the cross-validation AUC was calculated for each of the 20 imputed datasets and Rubin's rules were used to combine results across the imputed datasets.